Structural Analysis of Nonlinear Pricing

Yao Luo

University of Toronto

Isabelle Perrigne

Rice University

Quang Vuong

New York University

This paper proposes a new methodology for analyzing nonlinear pricing data. We establish identification of the model primitives with a known tariff and characterize the model restrictions on observables. We propose a quantile-based nonparametric estimator that achieves consistency at the parametric rate. We introduce unobserved product heterogeneity with an unknown tariff and show how our identification and estimation results extend. A Monte Carlo study analyzes the robustness of our methodology to menus of two-part tariffs. Analysis of cellular service data assesses the performance of various pricing strategies. We discuss extensions to network effects, multiproduct firms, bundling, differentiated products, and oligopolies.

I. Introduction

When facing heterogeneous consumers, a firm's prevalent strategy is to discriminate consumers by offering different prices across purchase sizes

This paper has been circulated under different titles. We thank the editor, three referees, Dan Ackerberg, Don Andrews, Gaurab Aryal, Steve Berry, Philippe Fevrier, Jin Hahn, Phil Haile, Ken Hendricks, Yunmi Kong, Guy Laroque, Pascal Lavergne, Vadim Marmer, Rosa Matzkin, Ariel Pakes, Daniel Quint, John Riley, Jean-Marc Robin, and Marc Rysman

Electronically published October 29, 2018 [Journal of Political Economy, 2018, vol. 126, no. 6] © 2018 by The University of Chicago. All rights reserved. 0022-3808/2018/12606-0008\$10.00

or qualities. This practice is called nonlinear pricing or second-degree price discrimination and can be found in electricity, telecommunication, and advertising, among others. See Wilson (1993) for examples. Spence (1977), Mussa and Rosen (1978), and Maskin and Riley (1984) provide seminal nonlinear pricing models within an imperfect information framework in which firms do not know the consumers' tastes for their products, leading to adverse selection. The firm designs a tariff discriminating consumers while endogenizing the offered quantity/quality. This tariff is incentive compatible or revealing and requires giving up some rents to consumers. The resulting optimal price schedule is concave in quantity, implying discounts.¹

The economic importance of price discrimination has led to an abundant empirical literature, with early studies by Lott and Roberts (1991) and Shepard (1991) focusing on evidence of nonlinear pricing. More recently, empirical studies have evaluated the impact of nonlinear pricing on profits, consumer surplus, and economic efficiency using a random utility discrete choice model for consumers' preferences to recover the consumers' taste distribution treating the price schedule as exogenous (see Leslie 2004; McManus 2007; Cohen 2008; Economides, Seim, and Viard 2008). A third trend has endogenized the price and quantity schedules to estimate the demand and cost structure (see Ivaldi and Martimort 1994; Miravete 2002; Crawford and Shum 2007). The latter use Mussa and Rosen (1978) to measure quality distortion in the cable TV industry. They consider a finite number of types/tastes with an equal number of of-

for constructive comments. We also benefited from participants' comments at the Stanford Institute for Theoretical Economics, Cowles Summer Conference, Econometric Society European Meeting, Latin American Meeting of the Econometric Society, Institute D'Économie Industrielle Workshop on the Econometrics of Industrial Organization, North American Summer Meeting of the Econometric Society and seminars at California Institute of Technology, Carnegie Mellon University, Massachusetts Institute of Technology, New York University, Paris School of Economics, Simon Fraser University, University of Arizona, University of California at Davis, University of California at Los Angeles, University College London, University of Fortaleza, University of Maryland, University of Montreal, University of Southern California, and University of Wisconsin-Madison. Luo thanks the Networks, Electronic Commerce, and Telecommunications Institute (www.NETinst.org) and Social Science and Humanities Research Council grant 430-2016-00077 for financial support, while Perrigne and Vuong gratefully acknowledge financial support from the National Science Foundation through grant SES-1148149. Perrigne thanks the Department of Economics at New York University for financial support while revising this paper. Data are provided as supplementary material online.

¹ Extensions to oligopoly or multiple products include Oren, Smith, and Wilson (1983), Ivaldi and Martimort (1994), Stole (1995, 2007), Armstrong (1996), Rochet and Choné (1998), Armstrong and Vickers (2001), Rochet and Stole (2003), and Bonatti (2011). Because of multidimensional screening, the optimal tariff is less tractable, though closed-form solutions might be obtained for some specifications.

² Other studies by Borenstein (1991), Borenstein and Rose (1994), and Busse and Rysman (2005) document the impact of competition on patterns of nonlinear pricing.

fered qualities and hence prices with known consumers' utility and firm's cost. Thus the observed market shares straightforwardly identify the frequencies of types. Using the first-order conditions for the consumer and the firm, the authors recover the unobserved qualities and types.

Our paper proposes a new methodology for the structural analysis of nonlinear pricing data when types are continuously distributed and the consumers' utility and firm's cost are unknown to the analyst. We consider Maskin and Riley's (1984) model with consumer exclusion. The model contains the main features for developing a general structural framework for nonlinear pricing data. The conclusion discusses how to extend our results to advanced pricing models with, for example, bundling, differentiated products, and competition. In the spirit of the empirical auction literature, we investigate the nonparametric identification of the model primitives from observables, which are the individual purchases and payments. See Laffont and Vuong (1996) and Athey and Haile (2007) for surveys.

We first consider a known firm's tariff. Our identification problem is reminiscent of those studied by Ekeland, Heckman, and Nesheim (2004) and Heckman, Matzkin, and Nesheim (2010), who study the nonparametric identification of hedonic price models. Instead of relying on instruments, we use the first-order conditions of both the consumer and the firm to identify the model primitives and exploit the one-to-one mapping between the unobserved consumer's type and his observed purchase. Given a cost parameterization, the consumer exclusion condition identifies its parameters. We then characterize the restrictions imposed by the nonlinear pricing model on observables, the main restriction being the concavity of the tariff. Next, we propose a computationally convenient nonparametric procedure for estimating the marginal payoff and the type distribution relying on empirical processes and quantiles. In contrast to Guerre, Perrigne, and Vuong (2000), our estimator is one-step and converges at the parametric rate and hence is less demanding in terms of sample size.

We then consider an unknown tariff with product unobserved heterogeneity. Such heterogeneity is important as observed payments and quantities are not perfectly related. This leads to a transformation model that identifies the tariff using Horowitz (1996) and Ekeland et al. (2004). We then extend our estimators and show how the estimated tariff affects their asymptotic distributions. We assess their properties via a Monte Carlo study and find that the estimation of the marginal cost needs an upper boundary correction. We also study the robustness of our approach to a menu of two-part tariffs as firms may face limits on tariff choices. Our methodology is robust when observations arise from two-part tariffs as long as the difference in slopes between two subsequent tariffs is not too important and/or the number of tariffs is not too small. Our findings

also highlight that a menu of two-part tariffs entails some ranges of quantities not offered by the firm.

The analysis of cellular service data shows the importance of product unobserved heterogeneity and supports the nonlinear pricing model. Counterfactuals assess the performance of nonlinear pricing relative to alternative pricing strategies such as two-part tariffs, minimum purchase, and quantity forcing in terms of profit and consumer surplus. Though the loss in a firm's profit might be small, these strategies exclude more consumers and those with low tastes are most hurt. The conclusion discusses extensions of our methodology opening several avenues for future research. These include network effects, multiproduct firms, bundling, product differentiation, and competition. Beyond nonlinear pricing, we can use our methodology to analyze pricing data with regulated firms and contract data in (say) environmental regulation, retailing, and labor, where incomplete information plays a key role.

The paper is organized as follows. Section II introduces the model, establishes identification with a known tariff, and characterizes the model restrictions. Section III develops a one-step quantile-based nonparametric estimator. It then introduces product unobserved heterogeneity with an unknown tariff and shows how our estimator extends. A Monte Carlo analysis assesses the properties of our estimators as well as a robustness analysis of our approach to a menu of two-part tariffs. Section IV presents the empirical analysis of cellular service data. Section V presents conclusions and discusses future lines of research. Appendix A provides the proofs of Section II, asymptotic properties of our estimators in Section III, and estimation of the tariff. Proofs of asymptotic properties are in online appendixes B and C.

II. Model, Identification, and Restrictions

A. The Model

1. Assumptions and Model Primitives

We consider the nonlinear pricing model by Maskin and Riley (1984), which is also one of the benchmark adverse selection models. This model contains the main components for analyzing nonlinear pricing data. Section V discusses extensions to network effects, bundling, multiple products, product differentiation, and competition. Each consumer is characterized by an adverse selection parameter θ capturing his taste or willingness to pay for the product. The type θ is the consumer's private information, while the firm knows only its distribution $F(\cdot)$. In general, the firm finds it unprofitable to serve those with low types. Specifically, the firm chooses optimally a threshold θ^* below which consumers are ex-

cluded from consumption. On the other hand, consumers with $\theta \ge \theta^*$ buy a positive quantity of the product.

Each consumer has a utility from consumption $U(Q;\theta) = \theta U_0(Q)$ and faces a tariff T(Q), where Q is the consumed quantity/quality. The function $U_0(\cdot)$ is the base utility.³ Thus, the consumer's net payoff from consumption is

$$\theta U_0(Q) - T(Q). \tag{1}$$

The firm incurs a cost C(Q) for producing Q for each consumer.⁴ Hereafter $[U_0(\cdot), F(\cdot), C(\cdot)]$ are the model primitives. The next assumptions follow the theoretical literature. A variable as a subscript indicates the derivative of a function with respect to this variable.

Assumption A1.

- i. The base utility function $U_0(\cdot)$ is twice continuously differentiable with $U_0(\cdot) \ge 0$, $U_{00}(\cdot) > 0$, and $U_{000}(\cdot) < 0$ on $[0, +\infty)$.
- ii. The type distribution $F(\cdot)$ is twice continuously differentiable with a density $f(\cdot) > 0$ on its support $[\underline{\theta}, \overline{\theta}]$, with $0 < \underline{\theta} < \overline{\theta} < \infty$. Moreover, $\rho(\theta) \equiv \theta \{[1 F(\theta)]/f(\theta)\}$ and its derivative are strictly positive for $\theta \in [\underline{\theta}, \overline{\theta}]$.
- iii. The cost function $C(\cdot)$ is twice continuously differentiable on $[0, +\infty)$ with marginal cost satisfying $C_Q(\cdot) > 0$ and $C_{QQ}(\cdot) \geq 0$ on $[0, \infty)$.

Assumptions A1-i, iii are standard, while A1-ii is specific to models of asymmetric information. The condition that $\rho(\cdot)$ is strictly positive and increasing guarantees sorting at equilibrium, which is an important property for solving the model; see also note 8.

The Optimization Problem and First-Order Conditions

The θ -consumer chooses a quantity/quality $Q(\cdot)$ as a function of his type. This quantity maximizes (1) and, provided the consumer chooses to buy, solves the first-order condition (FOC)

$$T_{\mathcal{Q}}(Q(\theta)) = \theta U_{0\mathcal{Q}}(Q(\theta)). \tag{2}$$

In words, his marginal utility must equal the marginal price at the quantity Q. In addition, consuming should provide a larger utility than not consuming, namely,

³ The multiplicative separability in θ is standard in the theoretical literature. It is an identifying assumption as a general functional form $U(Q, \theta)$ is not identified. An equivalent specification is $\int_0^Q \theta v_0(x) dx$, where $v_0(\cdot)$ is the consumer's willingness to pay for the marginal unit of the product, also called the inverse demand.

⁴ Previous versions of the paper consider a general cost function $C(\cdot)$ for the total amount produced across consumers. See Riley (2012) and n. 12.

$$\theta U_0(Q(\theta)) - T(Q(\theta)) \ge \theta U_0(0). \tag{3}$$

Equations (2) and (3) are the so-called incentive compatibility (IC) and individual rationality (IR) constraints.⁵

The firm's (expected) profit defined as the revenue from an agent consuming Q minus the cost for producing Q is

$$\int_{\theta^*}^{\bar{\theta}} [T(Q(\theta)) - C(Q(\theta))] f(\theta) d\theta.$$

Because the consumer's type θ is unknown to the firm, the profit is taken in expectation. The firm then chooses optimally the cutoff type θ^* , the offered quantity $Q(\cdot)$, and the tariff $T(\cdot)$ to maximize its profit subject to the IR and IC constraints.⁶ This gives

$$\max_{\theta^*, Q(\cdot), T(\cdot)} \int_{\theta^*}^{\bar{\theta}} [T(Q(\theta)) - C(Q(\theta))] f(\theta) d\theta, \tag{4}$$

subject to the IC and IR constraints. The next proposition gives the necessary conditions for the solution $[\theta^*, Q(\cdot), T(\cdot)]$. Let

$$M(\theta) \equiv \theta U_0(Q(\theta)) - C(Q(\theta)) - \frac{1 - F(\theta)}{f(\theta)} U_0(Q(\theta)).$$

Proposition 1. Suppose that the firm does not shut down.⁷ Under A1, there exists a unique threshold $\theta^* \in [\underline{\theta}, \bar{\theta})$ below which consumers are not served, that is, Q = 0. In addition, for $\theta \in [\theta^*, \bar{\theta}]$, the functions $Q(\cdot)$ and $T(\cdot)$ are nonnegative, once and twice continuously differentiable, respectively, with $Q_{\theta}(\cdot) > 0$, $T_{Q}(\cdot) > 0$, and $T_{QQ}(\cdot) < 0$ on $[\theta^*, \bar{\theta}]$ and $[Q(\theta^*), Q(\bar{\theta})]$. These functions solve the FOCs of the firm's optimization problem (4), namely,

$$\theta U_{0Q}(Q(\theta)) = C_Q(Q(\theta)) + \frac{1 - F(\theta)}{f(\theta)} U_{0Q}(Q(\theta))$$
 (5)

and

- ⁵ Despite the right-hand side of (3) depending on θ , there are no countervailing incentives because (3) is equivalent to $\theta[U_0(Q) U_0(0)] T(Q) \ge 0$, which is strictly increasing in θ by A1-i for any given Q. See Lewis and Sappington (1989) and Maggi and Rodriguez-Clare (1995) for countervailing incentives.
- ⁶ If the firm can discriminate consumers on the basis of some observed characteristics such as their age or location, known as third-degree price discrimination, these characteristics appear in $F(\cdot)$ as conditioning variables and/or in $U_0(\cdot)$ as additional variables and hence in the resulting tariff T(Q).
- ⁷ Shutdown corresponds to $M(\bar{\theta})$ ≤ 0; i.e., it is not profitable to serve any consumer. As a matter of fact, the function $M(\cdot)$ is strictly increasing on $[0, \infty)$ under A1 and (5). Thus, there are three cases: $0 \le M(\underline{\theta})$, $M(\underline{\theta}) < 0 < M(\bar{\theta})$, and $M(\bar{\theta}) \le 0$. The first two cases correspond to no exclusion and consumer exclusion, respectively. Proposition 1 excludes the third case, where it is not profitable for the firm to operate.

$$T_{Q}(Q(\theta)) = \theta U_{0Q}(Q(\theta)), \tag{6}$$

for $\theta \in [\theta^*, \overline{\theta}]$. If $M(\underline{\theta}) < 0$, then $\theta^* \in (\underline{\theta}, \overline{\theta})$ and θ^* solves the optimal exclusion condition

$$M(\theta^*) \equiv \theta^* U_0(Q(\theta^*)) - C(Q(\theta^*)) - \frac{1 - F(\theta^*)}{f(\theta^*)} U_0(Q(\theta^*)) = 0.$$
 (7)

If $M(\underline{\theta}) \geq 0$, then $\theta^* = \underline{\theta}$ and all consumers are served. In either case, the θ^* -consumer receives zero net utility; that is, the boundary condition $\theta^* U_0(Q(\theta^*)) = T(Q(\theta^*))$ holds.

For formal proofs of proposition 1, see Maskin and Riley (1984) or Tirole (1988). Equation (5) says that the marginal payoff for each type equals the marginal cost plus a nonnegative distortion term due to incomplete information. Hence, all consumers buy less than the efficient (first-best) quantity except for the $\bar{\theta}$ -consumer for whom there is no distortion. Once $Q(\cdot)$ is obtained from (5), (6) characterizes the optimal tariff $T(\cdot)$ using the boundary condition. Equation (7) expresses the trade-off between expanding the customer base and lowering the tariff. The boundary condition says that the agent with the cutoff type θ^* receives no rent. Under A1, the second-order condition of the consumer's problem is satisfied with a strictly increasing and concave tariff. In addition, differentiating (5) with respect to θ leads to a strictly increasing quantity schedule.⁸

B. Identification

This section establishes identification of the model when the tariff $T(\cdot)$ is known. Section III.B introduces product unobserved heterogeneity and relaxes this assumption. The model primitives $[U_0(\cdot), F(\cdot), C(\cdot)]$ are the consumer's base utility, his type distribution, and the firm's cost function. In addition to $T(\cdot)$, the analyst observes consumers' purchased quantities from a single market. Thus, identification investigates whether the primitives can be uniquely recovered from the observables $[T(\cdot), G^{Q^*}(\cdot)]$, where $G^{Q^*}(\cdot)$ is the quantity distribution.

$$\rho_{\theta}(\theta)U_{0Q}(Q(\theta)) = Q_{\theta}(\theta)\big[C_{QQ}(Q(\theta)) - \rho(\theta)U_{0QQ}(Q(\theta))\big].$$

Thus under A1, $Q_{\theta}(\cdot) \ge 0$. This property guarantees sorting at equilibrium and profit maximization.

 9 The superscript * refers to a truncation as only consumers with a type $\theta \geq \theta^{*}$ consume the good. As is common, the analyst does not have information on consumers who choose the outside option. Otherwise, we could identify the proportion of such consumers $F(\theta^{*})$ and hence $F(\cdot)$ instead of $F^{*}(\cdot)$ on $[\theta^{*},\bar{\theta}]$ from proposition 2. The data may also provide some exogenous agent's characteristics X such as his age, income, etc. Because second degree price discrimination imposes the same tariff across consumers, we can view θ as aggregating a consumer's observed and unobserved heterogeneity. See n. 20 and Luo, Perrigne, and Vuong (2015) for the introduction of market and consumer heterogeneity in a more general setting.

⁸ Specifically, we obtain

1. Discussion

An FOC similar to (6) arises in hedonic models studied by Ekeland et al. (2004) and Heckman et al. (2010), who show that the marginal utility is nonidentified without further restrictions. Without considering the firm's optimization problem, Ekeland et al. (2004) establish identification of the consumer utility and the distribution of unobserved heterogeneity up to location and scale by exploiting variations in some consumers' continuous exogenous variables that are independent of the term of unobserved heterogeneity. Heckman et al. (2010) also consider exogenous variables for single and multimarket data. Given that a general utility function $U(Q, \theta)$ is not identified, the former paper considers separable additivity in θ , while the latter considers alternative functional forms. Taking the logarithm of (6) gives $\log T_0(Q) = \log U_{00}(Q) + \log \theta$. In a standard setup, $U_{0Q}(\cdot)$ would be identified by assuming $E[\log \theta|Q] = 0$. But from (5), $Q = Q(\theta)$, creating an endogeneity problem traditionally solved with instruments. In nonlinear pricing, it seems difficult to find instruments that are correlated with Q but independent of θ . Consequently, the consumer's FOC (6) is not sufficient to identify the utility function and the type distribution.¹⁰

Our problem is also reminiscent of identification in incomplete information models that lead to an equilibrium monotonic relationship between an observable and the agent's private information. In auctions, Guerre et al. (2000) exploit this one-to-one mapping to recover the bidders' private value distribution. When the model contains more primitives to identify, several strategies can be entertained. When bidders are risk averse, Guerre, Perrigne, and Vuong (2009) exploit exogenous variations in the number of bidders leading to some exclusion restrictions to identify the bidders' utility function. In contract models, recent papers use exogenous variations and exclusion restrictions to identify the model primitives. D'Haultfoeuille and Février (2016) exploit exogenous variations in labor contracts that do not affect the agents' primitives to identify the latter. An alternative strategy is to exploit FOCs for both the principal and the agent to identify the model primitives as in Perrigne and Vuong (2011) for a procurement model with adverse selection and moral hazard. We follow this strategy and exploit (5) and (6). Our identification results do not require the existence of exogenous variables characterizing the agent and/or principal.¹¹

¹⁰ Equation (6) can also be viewed as a nonseparable model $Y = s(\theta)$ with $Y = T_Q(Q)$ and $s(\theta) = \theta U_{0Q}[Q(\theta)]$. The function $s(\cdot)$ is invertible in θ because $T_Q(\cdot)$ is strictly decreasing while $Q(\cdot)$ is strictly increasing in θ from proposition 1. Chesher (2003) and Matzkin (2003) show identification of this model under restrictive assumptions such as a uniform distribution on [0, 1] for θ .

¹¹ Gandhi, Navarro, and Rivers (2017) use this approach to identify production functions.

2. Identification of $[U_0(\cdot), F(\cdot), C(\cdot)]$

We first note that a scale normalization is necessary since $U(Q;\theta) = \theta U_0(Q)$, where both the type θ and the base utility $U_0(\cdot)$ are unknown. It suffices to multiply θ by some positive constant and to divide $U_0(\cdot)$ by this constant to obtain observationally equivalent structures as formalized in lemma 1 in appendix A. On the other hand, there is no need for a location normalization as the tariff fixes the location of $\theta U_0(Q)$ through the boundary condition $\theta^* U_0(Q(\theta^*)) = T(Q(\theta^*))$. We propose a convenient scale normalization in assumption B2.

Considering an adverse selection model, D'Haultfoeuille and Février (2007) show that at least one of their three primitives, namely, the surplus, the type distribution, or the cost function, needs to be known to identify the model. Instead, we parameterize the cost function. To keep expressions simple, we choose a constant marginal cost specification though alternative functional forms such as $C(Q) = \kappa(1+Q)^{\gamma}$ can be entertained.¹²

Assumption B1. The cost function is of the form $C(Q) = \kappa + \gamma Q$, with $\kappa \ge 0$ and $\gamma > 0$, for $Q \ge 0$.

The model primitives become $[U_0(\cdot), F(\cdot), \kappa, \gamma]$ while the marginal cost $C_Q(Q(\theta))$ reduces to γ . Thus evaluating (5) and (6) at $\bar{\theta}$ and noting that $\bar{Q} = Q(\bar{\theta})$ since $Q(\cdot)$ is strictly increasing from proposition 1, γ is identified by $\gamma = T_O(\bar{Q})$.

We now turn to the identification of the marginal base utility $U_{0Q}(\cdot)$ and the unobserved type distribution $F(\cdot)$. The basic idea is to exploit the monotonicity of the equilibrium quantity schedule $Q(\cdot)$, which defines a one-to-one mapping between the unobserved consumer type θ and the observed quantity Q. Our argument is based on quantiles. Let $\theta(\alpha)$ and $Q(\alpha)$ denote the α -quantiles of the truncated type distribution $F^*(\cdot) = [F(\cdot) - F(\theta^*)]/[1 - F(\theta^*)]$ and the truncated consumption distribution $G^{2^*}(\cdot)$. We define $\theta(0) = \theta^*$ and $Q(0) = \underline{Q}$ so that $\theta(\cdot)$ and $Q(\cdot)$ are well defined on [0,1]. Rewriting (5) and (6) in terms of quantiles and noting that $[1 - F(\theta)]/f(\theta) = [1 - F^*(\theta)]/f^*(\theta)$, where $f^*(\cdot)$ is the density of $F^*(\cdot)$, lead to

$$heta(lpha)U_{0Q}(Q(lpha)) = \gamma + rac{1-lpha}{f^*(heta(lpha))}U_{0Q}(Q(lpha)),$$
 $T_Q(Q(lpha)) = heta(lpha)U_{0Q}(Q(lpha)),$

¹² Alternatively, we could consider the cost of the total amount produced, i.e., $C(\int_{\theta^*}^{\bar{\theta}}Q(\theta)f(\theta)d\theta)$. In this case, the marginal cost for the total amount produced is identified by combining (5) and (6) evaluated at the upper boundary $\bar{\theta}$. In the multiproduct case, Luo, Perrigne, and Vuong (2017) show that the variable cost is identified nonparametrically under some homogeneity assumption.

for $\alpha \in [0, 1]$. Using the relationship between the density and its quantile function, that is, $f^*(\theta(\alpha)) = 1/\theta_{\alpha}(\alpha)$, and $U_{0Q}(Q(\alpha)) = T_Q(Q(\alpha))/\theta(\alpha)$ in the first equation give

$$\frac{\theta_{\alpha}(\alpha)}{\theta(\alpha)} = \frac{T_{Q}(Q(\alpha)) - \gamma}{(1 - \alpha)T_{Q}(Q(\alpha))}.$$
 (8)

Integrating (8) from 0 to α and using $\theta(0) = \theta^*$ give

$$\log \frac{\theta(\alpha)}{\theta^*} = \int_0^\alpha \frac{1}{1-u} \left[1 - \frac{\gamma}{T_O(Q(u))} \right] du. \tag{9}$$

Thus, (9) expresses the quantile $\theta(\alpha)$ in terms of observables that are the marginal tariff $T_Q(\cdot)$ and the quantiles $Q(\alpha)$ of the quantity distribution $G^{Q^*}(\cdot)$ up to the knowledge of θ^* . It follows that $\theta(\cdot)$ is identified on [0, 1] up to θ^* , suggesting the following scale normalization.¹³

Assumption B2. $\theta^* = 1$.

Assumption B2 is a normalization as implied by lemma 1. With B2, the truncated type distribution is identified by the inverse of its quantile function, that is, $F^*(\cdot) = \theta^{-1}(\cdot)$, on $[\theta^*, \bar{\theta}]$. Using $T_Q(Q(\alpha)) = \theta(\alpha) U_{0Q}(Q(\alpha))$, the marginal base utility $U_{0Q}(Q(\alpha))$ is expressed in terms of observables that are the marginal tariff and the quantiles of the quantity distribution as $\theta(\alpha)$ can be replaced by (9). It follows that $U_{0Q}(Q(\cdot))$ is identified on [0, 1]. Since $U_{0Q}(Q) = U_{0Q} \circ Q[G^{2*}(Q)]$, $U_{0Q}(\cdot)$ is identified on $[\underline{Q}, \bar{Q}]$. It remains to identify the cost parameter κ . We use the exclusion condition (7) for this purpose. The next proposition formalizes the identification of $[U_0(\cdot), F^*(\cdot), \kappa, \gamma]$.

PROPOSITION 2. Under assumptions A1 and B1, when there is exclusion, the cost parameters are identified by

$$\gamma = T_{\mathcal{Q}}(\bar{\mathcal{Q}}), \quad \kappa = \gamma \left(\frac{T(\underline{\mathcal{Q}})}{T_{\mathcal{Q}}(\underline{\mathcal{Q}})} - \underline{\mathcal{Q}} \right).$$

Moreover, under B2, the base utility $U_0(\cdot)$ is identified on $[\underline{Q}, \overline{Q}]$ as $U_0(Q) = T(\underline{Q}) + \int_{\underline{Q}}^{\underline{Q}} U_{0Q}(x) dx$. The truncated consumers' type distribution $F^*(\cdot)$ is identified on $[\theta^*, \overline{\theta}]$.

The primitives $U_0(\cdot)$ and $F^*(\cdot)$ are not identified on $[0, \underline{Q})$ and $[\underline{\theta}, \theta^*)$, respectively. Intuitively, the purchase and price data do not provide any

¹³ There are alternative normalizations as long as they involve a value in $[\theta^*, \bar{\theta}]$. On the other hand, any normalization in $[\underline{\theta}, \theta^*]$ is useless as the type distribution is not identified on that range. Despite having 1-u in the denominator, the integral in (9) exists at $\alpha=1$ as shown in the proof of lemma 2.

variation to identify these functions on those ranges as the minimum observed quantity is \underline{Q} . In the case of no exclusion, that is, all agents consume a strictly positive quantity, the fixed cost κ is not point identified. We can, however, provide an upper bound for κ using $M(\underline{\theta}) \geq 0$. Namely, $\kappa < \overline{\kappa} \equiv \gamma[[T(Q)/T_Q(Q)] - Q]$.

C. Model Restrictions

We characterize the restrictions imposed by the model on observables. These restrictions can be used to test the model's validity. As in Section II.B, we assume that $T(\cdot)$ is known. In Section IV, we discuss restrictions introduced by unobserved product heterogeneity. For every consumer, we observe his payment and consumption (t, Q), where t = T(Q) and Q is endogenously determined. The next lemma provides necessary and sufficient conditions on the tariff $T(\cdot)$ and the consumers' purchased quantity distribution $G^{Q^*}(\cdot)$, to be rationalized by a nonlinear pricing structure $[U_0(\cdot), F(\cdot), \kappa, \gamma]$.

Lemma 2 (Rationalization lemma). A structure $[U_0(\cdot), F(\cdot), \kappa, \gamma]$ satisfying A1, B1, and B2 rationalizes $[T(\cdot), G^{Q^*}(\cdot)]$ if and only if the latter satisfies the following conditions:

- i. $G^{\varrho^*}(\cdot)$ is once continuously differentiable with density $g^{\varrho^*}(\cdot) > 0$ on a compact support $[Q, \bar{Q}] \subseteq \mathbb{R}_+$.
- ii. $T(\cdot)$ is twice continuously differentiable, strictly increasing, and concave; namely, $T_Q(\cdot) > 0$ and $T_{QQ}(\cdot) < 0$ on $[Q, \bar{Q}]$.
- iii. $T(Q) \ge QT_Q(Q)$.

Condition i indicates that $G^{Q^*}(\cdot)$ is less smooth than the type distribution $F(\cdot)$. This follows from (8). We note that a positive density $g^{Q^*}(\cdot)$ arises not only from $f(\cdot) > 0$ but also from the optimal concave tariff. As discussed in our robustness analysis of Section III.C, a vanishing density on some intervals corresponding to quantities not offered by the firm suggests that the tariff is a menu of two-part tariffs. Condition iii relates to the fact that there is no firm shutdown since we observe (t, Q). When there is exclusion, it follows from $\kappa \ge 0$ and proposition 2. Condition ii is the main restriction of the nonlinear pricing model. It requires that the tariff is strictly increasing and concave. As the proof shows, condition ii ensures most of the properties of the model such as $Q(\cdot)$ being strictly increasing in θ as well as the model assumptions on the utility, hazard rate, and cost in A1. This result corroborates the early reduced-form literature displaying empirical evidence of a concave tariff to justify nonlinear pricing. Lemma 2 shows that the model of incomplete information imposes a few additional restrictions that are relatively mild.

III. Estimation and Unobserved Heterogeneity

Our estimation method follows identification and relies on (8) and (9), which involve the marginal tariff. Section III.A considers estimation with a known tariff $T(\cdot)$. Section III.B addresses identification and estimation of $T(\cdot)$ with unobserved product heterogeneity. Section III.C provides Monte Carlo evidence and a robustness analysis to two-part tariffs. In contrast to the previous literature on estimation of incomplete information models (e.g., Guerre et al. 2000), we develop one-step estimators for both the utility function $U_0(\cdot)$ and the type density $f^*(\cdot)$. Our estimators achieve the parametric rate, thereby circumventing the large data requirements associated with nonparametric estimators.

A. A One-Step Nonparametric Procedure

Our new estimator is based on quantiles. Let N denote the number of consumers purchasing a quantity Q_i , $i=1,2,\ldots,N$. We first consider the estimation of the marginal cost γ based on the identifying relationship $\gamma=T_Q(\bar{Q})$ from proposition 2. We use a maximum estimator for \bar{Q} , namely, $Q_{\max}=\max_i Q_i$, which converges at a fast rate. As shown in Section III.C, we can slightly modify this estimator to improve its finite sample properties. Regarding the fixed cost κ , the identifying equation of proposition 2 involves Q, which we estimate by $Q_{\min}=\min_i Q_i$. This gives

$$\hat{\gamma} = T_Q(Q_{\text{max}}), \quad \hat{\kappa} = \hat{\gamma} \left(\frac{T(Q_{\text{min}})}{T_O(Q_{\text{min}})} - Q_{\text{min}} \right). \tag{10}$$

Next, in view of (8) and (9), our estimators for $\theta(\cdot)$ and its derivative $\theta_{\alpha}(\cdot)$ are

$$\hat{\theta}(\alpha) = \exp\left\{ \int_{0}^{\alpha} \frac{1}{1 - u} \left[1 - \frac{\hat{\gamma}}{T_{Q}[\hat{Q}(u)]} \right] du \right\},$$

$$\hat{\theta}_{\alpha}(\alpha) = \frac{\hat{\theta}(\alpha)}{1 - \alpha} \frac{T_{Q}[\hat{Q}(\alpha)] - \hat{\gamma}}{T_{Q}[\hat{Q}(\alpha)]},$$
(11)

for $\alpha \in [0,1]$, where $\hat{Q}(\cdot)$ is an estimator of the quantile function from the observed quantities. We discuss estimation of the quantile function later. The relationships $f^*[\theta(\alpha)] \equiv [f^* \cdot \theta](\alpha) = 1/\theta_\alpha(\alpha)$ and $U_{0Q}[Q(\alpha)] \equiv [U_{0Q} \cdot Q](\alpha) = T_Q[Q(\alpha)]/\theta(\alpha)$ then provide estimators for the type density and the marginal base utility, at their $\theta(\alpha)$ and $Q(\alpha)$ quantiles, respectively:

$$\widehat{f^* \circ \theta}(\alpha) = \frac{1}{\widehat{\theta}_{\alpha}(\alpha)}, \quad \widehat{U_{0Q} \circ Q}(\alpha) = \frac{T_{\mathcal{Q}}[\widehat{Q}(\alpha)]}{\widehat{\theta}(\alpha)}.$$
 (12)

Hence, estimators at any value $\theta \in [\theta^*, \hat{\theta}] = [\theta^*, \hat{\theta}(1)]$ and $Q \in [Q_{\min}, Q_{\max}]$ are

$$\hat{f}^*(\theta) = \widehat{f^* \cdot \theta}[\hat{\theta}^{-1}(\theta)], \quad \hat{U}_{0Q}(Q) = \widehat{U_{0Q} \cdot Q}[\hat{G}^{Q^*}(Q)], \tag{13}$$

where $\hat{\theta}^{-1}(\cdot)$ is the estimated inverse of $\hat{\theta}(\cdot)$ and $\hat{G}^{Q^*}(\cdot)$ is the empirical distribution of $\{Q_i, i = 1, ..., N\}$. Since $\theta^{-1}(\cdot) = F^*(\cdot), \hat{\theta}^{-1}(\cdot)$ estimates the truncated type distribution $F^*(\cdot)$.

To implement (13), we need an estimator $\hat{Q}(\cdot)$ of the quantile function $Q(\cdot)$. We use the (generalized) inverse $\hat{G}^{Q^*-1}(\cdot)$ of the empirical distribution $\hat{G}^{Q^*}(\cdot)$. In particular, $\hat{Q}(\cdot)$ is a left-continuous step function on (0,1] with steps at $1/N < 2/N < \cdots < (N-1)/N$ with values equal to the ordered statistics $Q^1 \equiv Q_{\min} < Q^2 < \cdots < Q^N \equiv Q_{\max}$. At $\alpha=0$, we define $\hat{Q}(0) = Q_{\min}$. Thus the integral in (11) reduces to the simple expression

$$\log \hat{\theta}(\alpha) = \sum_{j=1}^{J-1} \int_{(j-1)/N}^{j/N} \frac{1}{1-u} \left[1 - \frac{\hat{\gamma}}{T_{Q}(Q^{j})} \right] du$$

$$+ \int_{(J-1)/N}^{\alpha} \frac{1}{1-u} \left[1 - \frac{\hat{\gamma}}{T_{Q}(Q^{J})} \right] du$$

$$= \sum_{j=1}^{J-1} \left[1 - \frac{\hat{\gamma}}{T_{Q}(Q^{j})} \right] \log \left(\frac{N-j+1}{N-j} \right)$$

$$+ \left[1 - \frac{\hat{\gamma}}{T_{Q}(Q^{J})} \right] \log \left(\frac{N-J+1}{N(1-\alpha)} \right),$$
(14)

for $(J-1)/N \le \alpha \le J/N$, where $J=1,2,\ldots,N$. We remark that $\hat{\theta}(\alpha)$ is continuous and increasing in $\alpha \in [0,(N-1)/N]$ since $T_{\mathcal{Q}}[\hat{Q}(u)] > T_{\mathcal{Q}}[Q_{\max}] = \hat{\gamma}$ by concavity of $T(\cdot)$. ¹⁴ Thus, the inverse $\hat{\theta}^{-1}(\cdot)$ can be readily computed from (14) for any $\theta \in [1,\hat{\theta}[(N-1)/N]]$.

Up to now, we have considered that the analyst knows the payment schedule and has data on purchased quantities. When the analyst knows the payment schedule but has data on payments $\{t_i, i = 1, ..., N\}$, we can express our estimators in terms of the latter. Since $Q = T^{-1}(t)$ and $T_Q(Q) = 1/T_t^{-1}(t)$, then (10)–(13) become

$$\hat{\gamma} = \frac{1}{T_t^{-1}(t_{\text{max}})}, \quad \hat{\kappa} = \hat{\gamma}(t_{\text{min}}T_t^{-1}(t_{\text{min}}) - T^{-1}(t_{\text{min}})), \quad (15)$$

For $(N-1)/N < \alpha \le 1$, the last term of (14) vanishes since $T_{\mathcal{Q}}(\mathcal{Q}^N) = T_{\mathcal{Q}}(\mathcal{Q}_{\max}) = \hat{\gamma}$. Thus, $\hat{\theta}(\alpha) = \hat{\theta}[(N-1)/N]$, which is finite and constant for $\alpha \in [(N-1)/N, 1]$.

$$\hat{\theta}(\alpha) = \exp\left\{ \int_0^\alpha \frac{1}{1-u} \left[1 - \hat{\gamma} T_t^{-1}(\hat{t}(u)) \right] \right\} du,$$

$$\hat{\theta}_{\alpha}(\alpha) = \frac{\hat{\theta}(\alpha)}{1-\alpha} \left[1 - \hat{\gamma} T_t^{-1}(\hat{t}(\alpha)) \right],$$
(16)

$$\widehat{f^* \circ \theta}(\alpha) = \frac{1}{\widehat{\theta}_{\alpha}(\alpha)}, \quad \widehat{U_{0Q} \circ Q}(\alpha) = \frac{1}{\widehat{\theta}(\alpha)T_t^{-1}[\widehat{t}(\alpha)]}, \quad (17)$$

$$\hat{f}^*(\theta) = \widehat{f^* \cdot \theta} (\hat{\theta}^{-1}(\theta)), \quad \hat{U}_{0Q}(Q) = \widehat{U_{0Q} \cdot Q} [\hat{G}^{i*}(T(\cdot))], \quad (18)$$

where $t_{\min} = \min_i t_i$, $t_{\max} = \max_i t_i$, $\hat{t}(\cdot) = \hat{G}^{t^*-1}(\cdot)$ is the estimator of the quantile function $t(\cdot)$ of $G^{t^*}(\cdot)$ with $\hat{t}(0) \equiv \underline{t}$, and $\hat{G}^{t^*}(\cdot)$ is the empirical truncated payment distribution.

Section B of appendix A derives the consistency and asymptotic distributions of the above estimators under the following assumption on the data-generating process.

Assumption C1. The unobserved types θ_i , i = 1, ..., N, are independent and identically distributed (i.i.d.) as $F^*(\cdot)$.

Since $Q_i = Q(\theta_i)$, for i = 1, ..., N, the observed quantities are also i.i.d. In particular, we obtain \sqrt{N} -consistency of $\hat{U}_{0Q}(\cdot)$ and $\hat{f}^*(\cdot)$, which arises from (8) and (9). Specifically, (9) shows that $\theta(\cdot)$ can be estimated at the parametric rate by integrating the quantile process. Thus, its derivative $\theta_{\alpha}(\cdot)$ and the density $f^*(\cdot)$ can also be estimated at the parametric rate by (8). Achieving the parametric rate is new. It is useful in practice as it allows for the analysis of medium-size data sets while avoiding the curse of dimensionality typically associated with nonparametric estimators.

B. Unobserved Heterogeneity

We now consider the case in which the analyst does not know the payment schedule $T(\cdot)$ but observes instead the pairs of payments and quantities (t_i, q_i) , i = 1, ..., N. The nonlinear pricing model of Section II implies that the observed payments and quantities lie on the curve t = T(q) as both depend on the consumer type θ_i . In practice, the observed prices and quantities may not lie on a curve, thereby calling for an additional source of randomness ϵ_i beyond θ_i . Several reasons can be invoked to justify this addition. For instance, the analyst may not observe perfectly the consumption. In addition, the product may be horizontally and/or vertically differentiated in more than a single dimension.

1. Discussion

We view ϵ_i as unobserved product heterogeneity capturing product differentiation. In Section IV, data from a cellular service company include

only the quantity q_i of phone calls measured in minutes besides the payment t_i . As a matter of fact, the payment includes additional services such as roaming, voice mail, ring tones, and so forth that we do not observe. More generally, data do not always contain information on all the product attributes that make up the payment. We assume that the analyst observes the quantity q_i while the nonlinear pricing mechanism is based on a latent quantity Q_i , which is an unknown function of q_i and ϵ_i , that is, $Q_i = m(q_i, \epsilon_i)$. Hereafter, Q_i is called the contracted quantity and the payment is based on this unobserved contracted quantity. The tariff equation becomes $t_i = T(Q_i) = T[m(q_i, \epsilon_i)]$, where $T(\cdot)$ is strictly increasing and concave. Taking the inverse gives $T^{-1}(t_i) = m(q_i, \epsilon_i)$. This general model is not identified.

We assume multiplicative separability of $m(\cdot, \cdot)$ in ϵ leading to $m(q_i, \cdot)$ $\epsilon_i \equiv m_0(q_i)\epsilon_i$, where $m_0(\cdot)$ is unknown while q_i and ϵ_i are mutually independent. The function $m_0(\cdot)$ captures the unobserved product attributes that depend on q. Specifically, each product attribute such as messages, roaming, voice mail, and ring tones in our application can be decomposed into two parts, one dependent on q and the other independent of q. The flexible function $m_0(\cdot)$ collects all the first parts, while ϵ aggregates all the second parts. Thus $m_0(\cdot)$ and ϵ both incorporate product differentiation.¹⁵ Taking the logarithm gives the transformation model $\log T^{-1}(t_i) = \log m_0(q_i) + \log \epsilon_i$. Under some location and scale normalizations, Horowitz (1996) establishes the nonparametric identification of $T(\cdot)$ with $m_0(\cdot)$ parametrically specified. Extending this result, Ekeland et al. (2004) show the nonparametric identification of $T(\cdot)$ and $m_0(\cdot)$ while Chiappori et al. (2015) obtain a similar result under a weaker independence assumption. Because the contracted quantity depends on the consumer type, that is, $Q_i = Q(\theta_i) = m_0(q_i)\epsilon_i$, this allows for dependence between θ_i and (q_i, ϵ_i) .

Regarding estimation, the transformation model leads to a \sqrt{N} -consistent estimator of $T^{-1}(\cdot)$ and hence of $T(\cdot)$ as shown by Chiappori et al. (2015), while the estimator of $m_0(\cdot)$ inherits the usual nonparametric rate. There are broadly two estimation methods for $T^{-1}(\cdot)$ involving either integrated kernel estimators following Horowitz (1996) or rank estimators following Chen (2002). Estimation of our model primitives also involves the marginal tariff $T_Q(\cdot)$. The \sqrt{N} -consistency is then lost since $T_Q(\cdot)$ is estimated at a slower rate. To maintain the parametric rate for

¹⁵ Chiappori, Komunjer, and Kristensen (2015) propose a test of independence of q and ϵ . We can relax the independence assumption of q and ϵ to allow for some heteroscedasticity as in Khan, Shin, and Tamer (2011) and/or some correlation as in Chiappori et al. (2015) using some instrumental variables.

¹⁶ We also face additional problems in the implementation of Horowitz (1996) and Chen (2002) estimators. In our case, the normalization $\underline{\epsilon} = 1$ leads to estimating the lower envelope of the scatter plot of the observations (t_i , q_i). As such, the estimate of the tariff function would be sensitive to outliers.

estimating the model primitives, we parameterize the tariff function, namely, $T^{-1}(\cdot) = T^{-1}(\cdot;\beta)$ for $\beta \in \mathbb{R}^{\dim \beta}$. This leads to the semiparametric transformation model $\log T^{-1}(t_i;\beta) = \log m_0(q_i) + \log \epsilon_i$ as studied by Linton, Sperlich, and Van Keilegom (2008). This specification has three appealing features. First, it delivers \sqrt{N} -consistency estimation of $T_Q(\cdot)$ and hence of our model primitives. Second, it allows us to assess the effect of estimating $T(\cdot)$ on the asymptotic distributions of the latter. Third, we can impose monotonicity and convexity on $T^{-1}(\cdot)$, while maintaining flexibility on how the observed quantity q_i affects the contracted quantity $Q_i = m_0(q_i)\epsilon_i$.

2. Estimation

Our model with product unobserved heterogeneity for those who consume is

$$\log T^{-1}(t;\beta) = \log m_0(q) + \log \epsilon, \tag{19}$$

where $T^{-1}(\cdot; \beta)$ is strictly increasing and convex in t and $\beta \in \mathbb{R}^{\dim \beta}$. The term ϵ is independent of q with $\mathrm{E}(\log \epsilon) = 0$ and $\log \epsilon$ distributed as $F^{\epsilon}(\cdot)$ on $[\log \underline{\epsilon}, \log \overline{\epsilon}] \subseteq \mathbb{R}$ with density $f^{\epsilon}(\cdot) > 0$. Identification of $[\beta, m_0(\cdot), F^{\epsilon}(\cdot)]$ is ensured by the identification of the nonparametric transformation model under location and scale normalizations. Hereafter, we assume that the observations $(t_i, q_i), i = 1, \ldots, N$, are i.i.d. Thus, $(q_i, \epsilon_i), i = 1, \ldots, N$, are also i.i.d., implying $Q_i = m_0(q_i)\epsilon_i, i = 1, \ldots, N$, are i.i.d., which is consistent with C1.

Let $\hat{\beta}$ be a \sqrt{N} -asymptotically normal estimator of β , that is, $\sqrt{N}(\hat{\beta}-\beta) \xrightarrow{D} \mathcal{N} \sim \mathcal{N}(0,\Omega)$. Section IV uses Linton et al.'s (2008) minimum distance estimator based on the independence of q and ϵ . Once $T^{-1}(\cdot)$ is estimated by $\hat{T}^{-1}(\cdot) \equiv T^{-1}(\cdot;\hat{\beta})$, we use (15)–(18), which rely on observed payments $\{t_i; i=1,\ldots,N\}$ to minimize estimation errors. Let $\tilde{\gamma}, \tilde{\kappa}, \tilde{\theta}(\cdot), \hat{\theta}_{\alpha}(\cdot), \hat{f}^* \circ \hat{\theta}(\cdot), \overline{U_{0Q} \circ Q}(\cdot), \tilde{f}^*(\cdot), \text{ and } \tilde{U}_{0Q}(\cdot) \text{ denote our estimators. We have$

$$\tilde{\gamma} = \frac{1}{\hat{T}_t^{-1}(t_{\text{max}})}, \quad \tilde{\kappa} = \tilde{\gamma}(t_{\text{min}}\hat{T}_t^{-1}(t_{\text{min}}) - \hat{T}^{-1}(t_{\text{min}})),$$
 (20)

$$\tilde{\theta}(\alpha) = \exp\left\{ \int_{0}^{\alpha} \frac{1}{1-u} \left[1 - \tilde{\gamma} \hat{T}_{\iota}^{-1}(\hat{t}(u)) \right] \right\} du,$$

$$\tilde{\theta}_{\alpha}(\alpha) = \frac{\tilde{\theta}(\alpha)}{1-\alpha} \left[1 - \tilde{\gamma} \hat{T}_{\iota}^{-1}(\hat{t}(\alpha)) \right],$$
(21)

$$\widetilde{f^* \circ \theta}(\alpha) = \frac{1}{\tilde{\theta}_{\alpha}(\alpha)}, \quad \widetilde{U_{0Q} \circ Q}(\alpha) = \frac{1}{\tilde{\theta}(\alpha)\hat{T}_t^{-1}[\hat{t}(\alpha)]}, \tag{22}$$

$$\tilde{f}^*(\theta) = \widetilde{f^* \cdot \theta} [\tilde{\theta}^{-1}(\theta)], \quad \tilde{U}_{0Q}(Q) = \widetilde{U_{0Q} \cdot Q} [\hat{G}^{\prime *}(\hat{T}(\cdot))], \tag{23}$$

where $\hat{T}_t^{-1}(\cdot) \equiv T_t^{-1}(\cdot;\hat{\beta})$, $\hat{G}^{t*}(\cdot)$ is the empirical distribution of payments, and $\hat{t}(\cdot) = \hat{G}^{t*-1}(\cdot)$ is its quantile estimator with $\hat{t}(0) \equiv \underline{t}$. Our estimators are straightforward to compute given $\hat{\beta}$. Letting $t^1 \equiv t_{\min} < t^2 < \cdots < t^N \equiv t_{\max}$ be the payment ordered statistics, we have

$$\log \tilde{\theta}(\alpha) = \sum_{j=1}^{J-1} \left[1 - \tilde{\gamma} \hat{T}_t^{-1}(t^j) \right] \log \left(\frac{N - j + 1}{N - j} \right)$$
$$+ \left[1 - \tilde{\gamma} \hat{T}_t^{-1}(t^J) \right] \log \left(\frac{N - J + 1}{N(1 - \alpha)} \right)$$

for $(J-1)/N \le \alpha \le J/N$, where $J=1,2,\ldots,N$. Moreover, $\tilde{\theta}(\cdot)$ is continuous and increasing in $\alpha \in [0,(N-1)/N]$ since $\hat{T}_t^{-1}[t(u)] < \hat{T}_t^{-1}[t_{\max}] = \hat{\gamma}$ by convexity of $T^{-1}(\cdot)$. On the other hand, $\tilde{\theta}(\alpha) = \tilde{\theta}[(N-1)/N]$ for $\alpha \in [(N-1)/N,1]$ (see n. 14). Section C of appendix A derives the consistency and \sqrt{N} -asymptotic distributions of the estimators (20)–(23).

C. Monte Carlo and Robustness to Two-Part Tariffs

We conduct a Monte Carlo study to assess the behavior of our estimators in finite samples as well as the robustness of our methodology to a menu of two-part tariffs.

1. A Monte Carlo Study

We use the following design. The utility function is $U(Q;\theta) = \theta Q^{1/2}$. The type distribution is $F(\theta) = 1 - (2 - \theta)^2$ with density $f(\theta) = 4 - 2\theta$ and support [1,2]. Thus $\rho(\theta) = -1 + (3/2)\theta$. The cost is C(Q) = 5 + 0.05Q. In contrast to A1, the density is not strictly positive everywhere as it vanishes at $\bar{\theta} = 2$. Using (5), we obtain $Q(\theta) = (15\theta - 10)^2$. From (7), the cutoff type θ^* is 4/3, which excludes 56 percent of consumers. Thus, the range of quantities offered is [100, 400]. We then use (6) and the boundary condition $\theta^* U_0(Q(\theta^*)) = T(Q(\theta^*))$ to solve for the tariff giving $T(Q) = (Q + 20\sqrt{Q} + 100)/30$. Hereafter, we consider three sample sizes: N = 250, 1,000, 4,000, where the latter corresponds to our empirical application.

First, we consider a known tariff T(Q). We draw randomly N values of θ in [4/3,2] and compute the corresponding quantities $Q=Q(\theta)$. We implement our quantile-based estimator of Section III.A to estimate $[U_0(\cdot),f(\cdot)]$. We replicate this exercise 1,000 times. The density $f(\theta)=4-2\theta$ has a right vanishing tail, which may cause a significant downward bias in the estimation of Q_{\max} and then an upward bias in the estimation of

 γ and κ through (10).¹⁷ Table 1 gives the estimated cost parameters. We remark an upward bias for $\hat{\gamma}$ of order 1 percent for N=250, reducing to 0.2 percent for N=4,000. The bias for estimating κ ranges from 1.16 percent to 0.25 percent. Moreover, the confidence intervals do not contain the true value of (γ, κ) . This calls for a correction of the upper boundary estimator. We apply the nonparametric support estimator by Devroye and Wise (1980), which reduces to $Q_{\text{max}} + h_0$, where h_0 is a bandwidth. We use the rule of thumb formula for h_0 in the kernel estimation of the density of Q. Table 1 also presents the estimates of (γ, κ) with the upper boundary correction. We observe a smaller bias, which varies from 0 percent to 0.4 percent for $\hat{\gamma}$ and 0.15 percent to 0.5 percent for $\hat{\kappa}$. More importantly, all the confidence intervals now contain the true values of (γ, κ) . Figure 1 displays $\hat{f}^*(\cdot)$ and $\hat{U}_{00}(\cdot)$ for the three sample sizes with the upper boundary correction. For N = 250, the results already present an excellent fit, while for N = 4,000, it is almost perfect in terms of precision and bias. This shows the necessity of adding a correction at the upper boundary in the estimation of cost parameters in small sample sizes.

Second, we consider unobserved product heterogeneity and an unknown tariff. We set $m_0(\cdot)$ as the identity function leading to $\log T^{-1}(t) =$ $\log q + \log \epsilon \equiv \log Q$. The inverse tariff is $T^{-1}(t) = 100 + 30t - 100 =$ $20\sqrt{30}\sqrt{t}$ leading to $T^{-1}(t;\beta) = \beta_0 + \beta_1 t + \beta_2 \sqrt{t}$. Because the transformation model requires normalizations, we set $\beta_0 = 100$ and $\beta_1 = 30$. We then use Linton et al.'s (2008) estimator, which minimizes a criterion based on the independence of q and $\hat{\epsilon}(\beta_2)$, where the latter comes from the residual of the nonparametric kernel regression of $\log T^{-1}(t; \beta_2)$ on q. Because the distribution of q is very skewed, we apply a Box-Cox transformation to q when running this nonparametric regression. The rest of the estimation follows Section III.B. Figure 2 displays $\hat{f}^*(\cdot)$ and $\hat{U}_{00}(\cdot)$ for the three sample sizes with the upper boundary correction. There is a bias for N = 250 that disappears when N = 1,000. Overall for N = 1,000 and N = 1,0004,000, the results present a good fit in terms of precision and bias. Comparing with figure 1, the confidence intervals tend to be wider than when the tariff is known. This is expected as unobserved product heterogeneity introduces an additional noise in the estimation. Overall, our Monte Carlo simulations show the good behavior of our estimators in small samples.

$$Pr(Q \le x | Q \ge 100) = Pr(\theta \le 2/3 + \sqrt{x}/15 | \theta \ge 4/3).$$

Using $F(\theta) = 1 - (2 - \theta)^2$ gives $G^{Q^*}(x) = -(\sqrt{x} - 10)(\sqrt{x} - 30)/100$ so that the distribution of $\log Q$ is $G^{\log Q^*}(y) = -(e^{y/2} - 10)(e^{y/2} - 30)/100$ on the support [log 100, $\log 400$]. Using sieves, we approximate numerically the quantiles of $\log q$ and $\log \epsilon$ from the quantiles of $\log Q$ subject to the independence of ϵ and q, the range of $\log \epsilon$ being 0.4, and $\operatorname{E}[\log \epsilon] = 0$. We then draw randomly N values of $\log q$ and $\log \epsilon$, independently.

¹⁷ We thank a referee for this point.

¹⁸ Random draws require more attention. From $Q(\theta)=(15\theta-10)^2$, the truncated distribution $G^{2^s}(\cdot)$ is

	<u> </u>		
Mean	Standard Deviation ($\times 10^3$)	90% Confidence Interval	
b = 2: No Boundary Correction			
.0505	.2560	[.0501, .0510]	
		[5.0203, 5.1067]	
.0502	.1200	[.0501, .0505]	
5.0260	12.4000	[5.0088, 5.0490]	
.0501	.0640	[.0500, .0502]	
5.0127	6.4000	[5.0036, 5.0243]	
b = 2: With Boundary Correction			
.0500	.2430	[.0496, .0504]	
5.0075	26.6000	[4.9705, 5.0545]	
.0499	.1140	[.0497, .0501]	
4.9890	11.8000	[4.9724, 5.0109]	
.0498	.0613	[.0498, .0500]	
4.9849	6.2000	[4.9761, 4.9959]	
	.0505 5.0580 .0502 5.0260 .0501 5.0127 .0500 5.0075 .0499 4.9890 .0498	Mean Deviation (×10³) $b = 2$: No Boundary Corn .0505 .2560 5.0580 27.8000 .0502 .1200 5.0260 12.4000 .0501 .0640 5.0127 6.4000 $b = 2$: With Boundary Corn .0500 .2430 5.0075 26.6000 .0499 .1140 4.9890 11.8000 .0498 .0613	

TABLE 1 Monte Carlo Cost Parameters: $\gamma = 0.05, \kappa = 5$

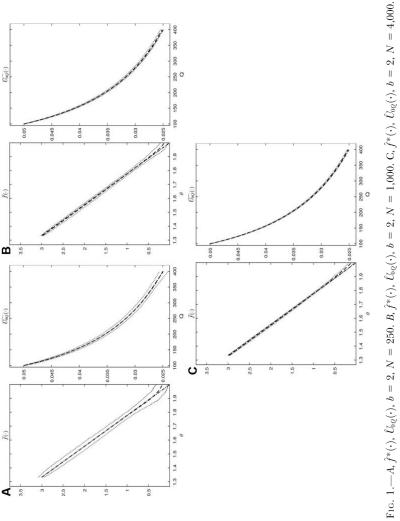
2. Robustness to Two-Part Tariffs

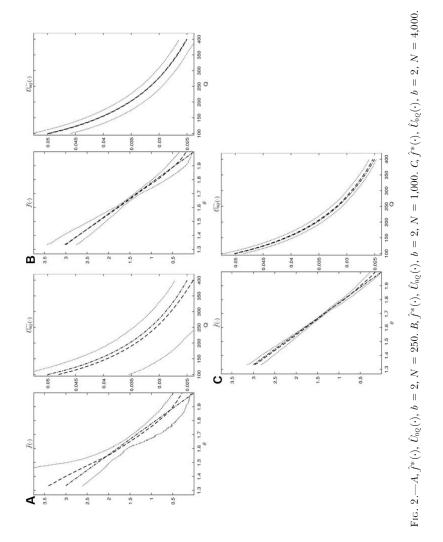
An interesting question is whether our methodology is robust when the firm faces limits on tariff choices and approximates the optimal tariff by a menu of two-part tariffs. Wong (2012) shows that a menu of two-part tariffs still leads to a quantity $Q(\cdot)$ monotonic in θ ; that is, there is no pooling, but some ranges of quantities are no longer offered/chosen, implying some discontinuities in $Q(\theta)$. This phenomenon amplifies when the number of two-part tariffs is small or when the difference in slopes of two adjacent tariffs is large. Thus not observing some ranges of quantities suggests that consumers face a menu of two-part tariffs. This contrasts with lemma 2-i, which requires the density $g^{Q*}(\cdot)$ to be strictly positive on the range $[Q, \overline{Q}]$ when the optimal tariff is offered.

We consider the same setting as before with a menu of M = 10 or 5 two-part tariffs. With a two-part tariff of the form $a_m + b_m Q$, this corresponds to 20 and 10 tariff parameters, respectively. These parameters are computed by solving

$$\min_{(a_n,b_n),m=1,...,M} \sum_{Q_\ell \in \mathcal{I}_n} \left[\frac{T(Q_\ell) - a_n - b_m Q_\ell}{T(Q_\ell)} \right]^2 + \left[\frac{T_Q(Q_\ell) - b_m}{T_Q(Q_\ell)} \right]^2,$$

where the Q_{ℓ} 's constitute a grid of $[Q, \bar{Q}]$ with a step size of 0.1 and the \mathcal{I}_m intervals partition $[Q, \bar{Q}]$. To simplify, we consider that there is no unobserved product heterogeneity; otherwise $\hat{T}(\cdot)$ would replace $T(\cdot)$ in the above minimization. Given the approximating menu of two-part tariffs, we then solve the consumer optimization problem for each draw $\theta_1, \ldots, \theta_N$, leading to new quantities Q_i , $i=1,\ldots,N$. We then apply our estimation





2543

method of Section III.A with boundary correction. It should be noted that the above approximation in two-part tariffs might differ from the menu of two-part tariffs that maximize the firm's profit. In practice, however, these two menus are very close as suggested by table 3 below.

Figures 3 and 4 display the estimated type density and marginal base utility function for three sample sizes when M=10 and 5. Not surprisingly, as M increases, our estimates improve. The observed "wavelets" are due to the absence of quantities in some intervals. Overall, when M=10, the estimation results for $U_{0Q}(\cdot)$ and $f^*(\cdot)$ offer an excellent fit to the truth. When M=5, $U_{0Q}(\cdot)$ remains well estimated while $\hat{f}^*(\cdot)$ would require additional smoothing. To summarize, our method is robust to a menu of two-part tariffs as long as their number is not too small or the underlying optimal tariff is not too concave as the gap in quantity values may render estimation of the model unsteady.

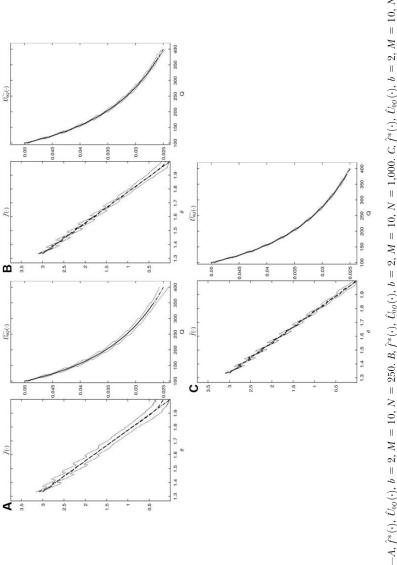
IV. Nonlinear Pricing in Mobile Service

This section analyzes mobile service data. Since we do not have information on all the components that make up the payment, we introduce product unobserved heterogeneity. Counterfactuals assess the benefits of nonlinear pricing over alternative pricing strategies.

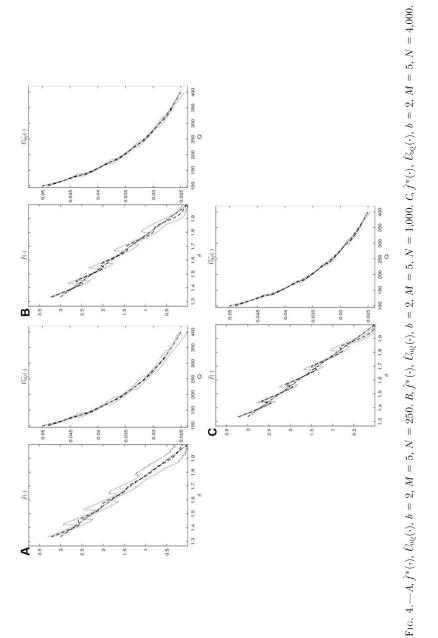
A. Data

We have voice data from a major mobile service company in a large Asian city. This company enjoys a market share of 72 percent in terms of consumers, 83 percent in terms of revenue, and 83 percent in terms of voice usage. Because of important price discounts on usage, consumers buy cellular phone services from a single company. This makes the monopoly assumption reasonable. A random sample of 4,000 consumers who are under the 2009 tariff covers the billing period of May 2009. For each consumer, we observe only the amount paid t and the total number of minutes consumed t. The first two rows of table 2 provide summary statistics on the number of minutes and the payment expressed in US dollars. Figure 5 displays the scatter plot of observations. A striking feature is that there are as many pairs (t, t, t) as observations. The data also show a large variability in prices at a given quantity and conversely. This arises from additional features that subscribers buy but that we do not observe.

¹⁹ The company changes its tariff every year. Consumers can switch to the new tariff upon request at no extra cost. In addition, consumers pay for what they consume, avoiding the typical problem associated with ex ante usage uncertainty. See Miravete (2002), Yao et al. (2012), and Grubb and Osborne (2015) for phone usage data and Ito (2014) for electricity usage data. Yao et al. (2012) find that the average ratio of usage is above 90 percent of allowance with a significant variability.



 $\text{Fig. 3.} - A, \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 250. \ B, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 1,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 4,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 4,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 2,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 2,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ b = 2, \ M = 10, \ N = 2,000. \ C, \\ \hat{f}^*(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot), \ \hat{U}_{0\mathcal{Q}}(\cdot$



This content downloaded from 128.042.192.005 on December 10, 2018 12:19:23 PM All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

	Observations	Mean	Median	Minimum	Maximum	Standard Deviation
\overline{t}	4,000	34.63	30.37	13.97	99.78	16.52
q	4,000	704.32	620.00	10	3,498.00	427.88
Q	4,000	1,650.15	1,087.00	195.20	9,171.03	1,561.51
$\hat{\theta}$	4,000	1.82	1.57	1.00	6.20	.82
$\hat{\epsilon}$	4,000	1.19	.94	.13	15.74	.96
Rent	4,000	23.37	12.68	0	213.20	29.31
Rent ratio	4,000	.49	.42	0	2.17	.38

TABLE 2 Summary Statistics

This includes roaming, ring tones, call forwarding, directory assistance, voice mail, and so forth. Regressing t_i on q_i and q_i^2 gives an R^2 of about 60 percent, confirming an important unobserved product heterogeneity while the mean tariff is strictly concave in q^{20} .

B. Estimation of the Tariff Function

In (19), we use a flexible spline-based specification for $T^{-1}(\cdot)$ of the form

$$T^{-1}(t;\beta) = \beta_0 + \beta_1 t + \sum_{k=1}^{K+2} \delta_k \psi_k(t), \tag{24}$$

where $\beta=(\beta_0,\beta_1,\delta_1,\ldots,\delta_{K+2})$ and K is the number of (interior) knots. The basis functions $\psi_k(\cdot)$ borrow from Dole (1999) and Meyer (2012). We estimate the inverse tariff imposing that it is increasing and convex. This is achieved with $\beta_1>0$, $\delta_k>0$, $k=1,\ldots,K+2$. As is well known, identification of the transformation model (19) requires two normalizations. We set $\hat{\beta}_0=-448.96$ and $\hat{\beta}_1=46.11$. To estimate $\delta=(\delta_1,\ldots,\delta_6)$, we use Linton et al.'s (2008) minimum distance estimator relying on independence. See Section D of appendix A for details. The estimated coefficients are $\hat{\delta}_1=3.171$, $\hat{\delta}_2=0.263$, $\hat{\delta}_3=1.502$, $\hat{\delta}_4=0.178$, $\hat{\delta}_5=0.022$, and $\hat{\delta}_6=0.000055$. To test $H_0:\hat{\delta}_1=\cdots=\hat{\delta}_6=0$, which corresponds to a linear tariff, we use a test statistic based on the sum of square errors (SSE), namely, (SSE ^c-SSE)/SSE, where the index c refers to the constrained model under H_0 . We find a value of 0.7542. The bootstrapped asymptotic distribution gives a p-value of .08, thereby rejecting the linear tariff at the 10 percent significance level.

The estimated tariff $\hat{T}(\cdot)$ as a function of $Q = \hat{m}_0(q)\hat{\epsilon}$ is displayed in figure 6. The tariff shows an important curvature as suggested by the above

Our data contain neither information on the amount paid specific to q nor observed consumer characteristics X. If data contain such X variables, we can include them by viewing θ as aggregating consumers' observed and unobserved heterogeneity, i.e., $\theta = r(X, \eta)$; see Luo et al. (2015).

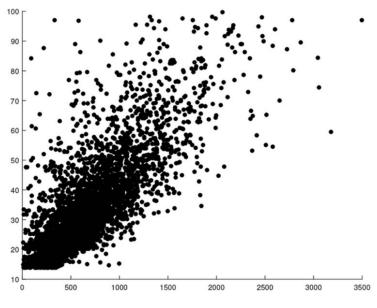
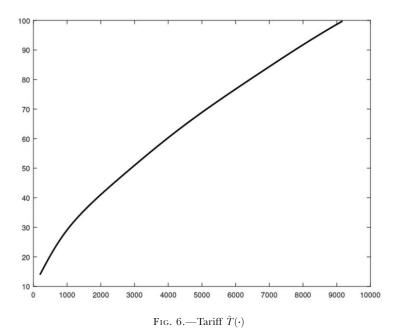


Fig. 5.—Scatter plot (q_i, t_i)

data analysis and estimated coefficients. Though not shown, $\hat{m}_0(\cdot)$ is increasing in q and convex with some irregularities for large values of q. This agrees with the economic intuition that the latent contracted quantity Q is increasing in voice usage q. Figure 7 displays a skewed density $\hat{f}^{\epsilon}(\cdot)$. Most of its values range in the [0,2] interval. From table 2, the variability in unobserved product heterogeneity $\hat{\epsilon}$ is more important than the variability in consumption of minutes as measured by their respective coefficients of variation, 0.81 and 0.61. To test the transformation model, we apply Chiappori et al.'s (2015) test of independence of q and ϵ . We follow their procedure and find a test statistic of 183.53 corresponding to a p-value of .242, thereby failing to reject the null assumption of independence. This corroborates our interpretation of $m_0(\cdot)$, which captures the unobserved product attributes dependent on q, while the term ϵ aggregates the independent ones. See Section D of appendix A.

Finally, we verify whether the restrictions imposed by the model in lemma 2 are satisfied. The estimated tariff is strictly concave in the contracted quantity Q, thereby satisfying condition ii. Regarding condition i, figure 8 displays the estimated density $\hat{g}^{Q*}(\cdot)$ of $\hat{Q}_j = T^{-1}(t_i;\hat{\beta})$. This density is strictly positive on $[Q_{\min}, Q_{\max}]$. From Section III.C, this supports a nonlinear tariff relative to a menu of two-part tariffs. In addition, condition iii is satisfied as 13.97 > 195.2/46.11 since $T(Q_{\min}) = 13.97$, $Q_{\min} = 195.2$, and $T_Q(Q_{\min}) = 1/T_i^{-1}(t_{\min}) = 1/46.11$. Thus, the nonlinear pricing model rationalizes the data.



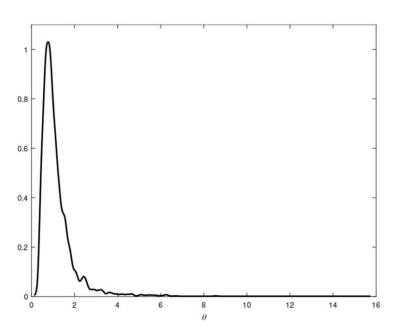


Fig. 7.—Density of unobserved heterogeneity $\hat{f}_{\epsilon}(\cdot)$

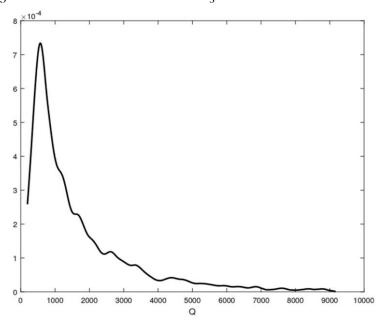


Fig. 8.—Density of contracted quantity $\hat{g}^{Q^*}(\cdot)$

C. Empirical Results

Using the upper boundary correction, we obtain $Q_{\text{max}} = 9,486.12$, leading to $\hat{\gamma} = 0.0063$. For the median consumer, this gives a variable cost of $0.0063 \times 1,087 = 6.85$ compared to a payment of 30.37. The estimated fixed cost is $\hat{\kappa} = 2.83$, representing 9.3 percent of the bill for the median consumer. Using its asymptotic distribution, the 95 percent confidence interval for $\hat{\gamma}$ is [0.00623, 0.00637], supporting the significance of $\hat{\gamma}$. Implementing the quantile estimator of Section III.C, figure 9 displays $\hat{\theta}(\cdot)$, which is strictly increasing, while figure 10 shows that $\hat{U}_{0Q}(\cdot)$ satisfies A1-i. Figure 11 displays $\hat{f}^*(\cdot)$. Because of the truncation at $\theta^* = 1$, we cannot identify the type density below θ^* . Fitting a beta distribution $\mathcal{B}(a,b)$ on estimated rescaled consumer types $x_i = (\hat{\theta}_i - 1)/(\theta_{\text{max}} - 1)$ by maximum likelihood gives the estimates $\hat{a} = 0.676$ and $\hat{b} = 3.298$. Figure 12 shows that $\hat{f}^*(\cdot)$ verifies A1-ii as $\theta = [1 - \hat{F}(\theta)]/\hat{f}(\theta)$ is strictly increasing. Finally, table 2 provides summary statistics on the estimated types and the informational rents left to consumers. Using (1), the individual informational rent is estimated by $\hat{\theta}_i \hat{U}_0(\hat{Q}_i) - \hat{T}(\hat{Q}_i)$, where $\hat{U}_0(\cdot) = \hat{T}(Q_{\min}) +$ $\int_{O_{-}} \hat{U}_{0Q}(q) dq$ and $\hat{\theta}_i = \hat{\theta}(\hat{\alpha}_i)$, with $\hat{\alpha}_i$ corresponding to the quantile of the payment t_i. The consumer's rent ratio (rent divided by payment) is, on average, 49 percent with an important variability arising from consumer

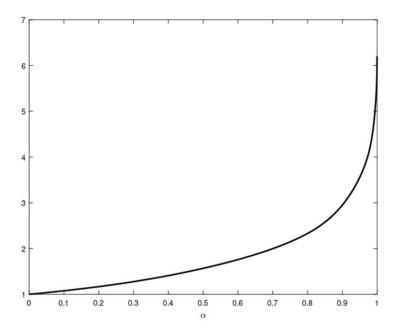


Fig. 9.—Quantile type $\hat{\theta}(\cdot)$

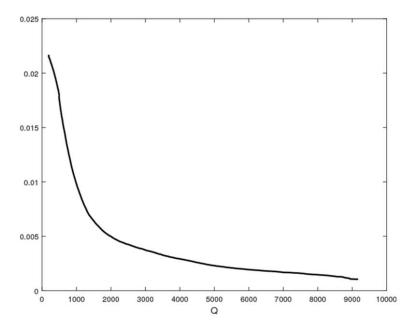


Fig. 10.—Marginal base utility $\hat{\it U}_{0\it Q}(\cdot)$

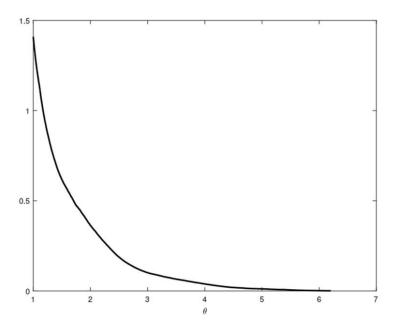


Fig. 11.—Type density $\hat{f}^*(\cdot)$

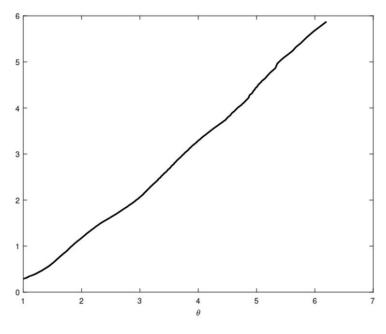


Fig. 12.— $\theta = [1 - \hat{F}(\theta)]/\hat{f}(\theta)$

heterogeneity. Because of skewness, we report its median, which is 42 percent.

D. Counterfactuals

We simulate the outcomes of three alternative pricing strategies that could be imposed by a regulatory agency. These counterfactuals assess the loss/gain associated with pricing strategies that are used by cellular service providers throughout the world. For each pricing strategy, the company can choose four price parameters that maximize its profit. We then measure the profit loss as well as the potential gain/loss for consumers. Since we cannot identify the type distribution and the utility function on their full supports, a natural question is how this influences our counterfactual results. Wong (2012) shows that it is generally more profitable for the firm to exclude more consumers when implementing alternative pricing strategies. This result relates to the dominance of nonlinear pricing in terms of a firm's profit for a large class of type densities. Thus, the nonidentification of $f(\theta)$ for values below the type cutoff θ^* does not constitute an obstacle to perform counterfactuals. In fact, we show that the alternative pricing strategies exclude more consumers. Because some consumers may consume less than Q_{\min} , we approximate the marginal base utility function for values below Q_{\min} by a linear function continuous at Q_{\min} with $U_0(0) = 0$.

Our counterfactuals are summarized in table 3, which provides the consumer surplus (informational rent), the firm's profit, the total welfare (as the sum of the two), the total contracted quantity, and the total payment. The nonlinear pricing column reports the actual values while the other columns correspond to the three alternative pricing strategies as proportions of the nonlinear pricing values. Table 4 assesses the winners and losers upon dividing the sample into four equal subsamples from the lowest to the highest consumer types.

A first counterfactual consists in a menu of two-part tariffs $t = c_j + p_j Q$, j = 1, 2, where c_j and p_j are the fixed fee and the marginal price. The values maximizing the firm's profit are $c_1 = 12.21$, $p_1 = 0.0168$, $c_2 = 34.65$,

 $\begin{tabular}{ll} TABLE~3\\ Comparisons~with~Alternative~Pricing~Strategies\\ \end{tabular}$

Indicators	Nonlinear Pricing	Two-Part Tariffs	Minimal Quantities	Plans
Consumer surplus	93,494	1.0043	.9437	.9335
Profit	85,608	.9933	.9616	.9610
Welfare	179,100	.9991	.9522	.9467
Total Q	6,600,600	.9995	.9909	.9828
Total payment	138,540	.9927	.9600	.9563
Consumers	4,000	.9630	.8345	.8228

TABLE 4				
Winners and Losers				

Group and Variable	Nonlinear Pricing	Two-Part Tariffs	Minimal Quantities	Plans
Low types:				
Consumer surplus	1,768	.9474	.2185	.1661
Total payment	18,686	1.0354	.5666	.4981
Total Q	418,360	1.2704	.8915	.7923
Medium low types:				
Consumer surplus	7,896	1.0288	.8582	.8287
Total payment	25,968	.9885	1.2064	1.2318
Total Q	807,440	.9906	1.3666	1.4108
Medium high types:				
Consumer surplus	20,376	1.0005	.9385	.9334
Total payment	35,746	1.0193	.8780	.8949
Total Q	1,524,900	1.0759	.7263	.7470
High types:				
Consumer surplus	63,454	1.0040	.9762	.9680
Total payment	58,139	.9645	1.0269	1.0182
Total Q	3,849,800	.9416	1.0277	1.0072

and $p_2 = 0.0093$. The results confirm that a simple two-part tariff does almost as well as nonlinear pricing as the firm's profit is smaller by less than 1 percent despite excluding about 4 percent of our sample of 4,000 consumers. Thus, the low-type group is the most hurt because of consumers' exclusion, though those who buy tend to consume more. Overall, this policy benefits the two medium groups as the marginal price tends to be lower than under nonlinear pricing. The high-type group is indifferent, though the members consume less because they do not benefit as much from price discounts. To summarize, simple two-part tariffs can perform efficiently at the social cost of excluding more consumers. The latter effect can be reduced by increasing the number of two-part tariffs.

A second counterfactual consists in a menu of linear tariffs with minimum purchase $t = p_j Q$ for $m_j \le Q < m_{j+1}, j = 1, 2, m_3 = \infty$, where m_j and p_j are the minimum quantity and the marginal price. The values maximizing the firm's profit are $m_1 = 1,107$, $p_1 = 0.0283$, $m_2 = 3,992$, and $p_2 = 0.0150$. This pricing strategy excludes 17 percent of consumers as the minimum payment becomes 31.33 while the minimum purchased quantity increases by more than 500 percent from 195 as observed in the data to 1,107. This pricing strategy greatly hurts the low-type group as their surplus is only 22 percent of what it is under nonlinear pricing. All the other groups are hurt as well, though not as much, with the high-type group losing the least. Overall, all the indicators of interest decrease by 4–6 percent, hurting the consumers the most.

A third counterfactual consists in a menu of quantity forcing or plans with a fixed maximum number of minutes, that is, $t = T_j$ with Q_j for j = 1, 2, where T_j and Q_j are the monthly fee and allowance in minutes. The values maximizing the firm's profit are $T_1 = 31.99$, $Q_1 = 1,139$, $T_2 = 62.25$,

and $Q_2 = 4,191$. This pricing strategy is commonly used in several countries, though the degree of customization remains important in the United States and Canada. This excludes 18 percent of consumers and hurts the low-type group because the increase in payment is important. The other groups lose at different levels because they either pay more or consume less than under nonlinear pricing. The loss in consumer surplus for the high-type group is only 3 percent as they would be willing to consume more. Overall, this pricing does not provide as much profit and consumer surplus, which decrease by 4 percent and 7 percent, respectively. These adverse effects can be reduced by offering a third plan targeted to low-type consumers with a low payment or by offering prepaid plans.

To summarize, despite a coarse menu in the three alternative pricing strategies, the reduction in consumer surplus and firm profit is quite small overall (less than 10 percent). On the other hand, consumer exclusion is much higher and the low-type consumers are hurt the most. Of course, these findings are contingent on the estimated structure, which shows an important heterogeneity and skewness in consumer types.

V. Conclusion and Extensions

This paper studies the identification and estimation of the canonical nonlinear pricing model. Identification is achieved by exploiting the first-order conditions of the firm and the consumer and the one-to-one mapping between the unknown consumer's type and his observed consumption. We propose a new quantile-based nonparametric estimator for the consumer's utility function and type density that achieves \sqrt{N} -consistency, a rate that is new for such model primitives in incomplete information models. In addition, we introduce unobserved product heterogeneity with an unknown tariff. We study the robustness of our estimator to two-part tariffs. Our empirical analysis of cellular service data supports the nonlinear pricing model and counterfactuals assess the performance of alternative pricing strategies.

Our paper proposes a general methodology for second- and also thirddegree price discrimination as observed characteristics can be introduced in the tariff. It represents a step toward the identification and estimation of advanced models of nonlinear pricing as we discuss next. Indeed, they lead to equations similar to (5) and (6). Identification, estimation, and unobserved product heterogeneity need to be adapted accordingly.

A. Network Effects

In some applications, such as in advertising, there is a network effect among agents. Specifically, let $Q_h = \int_{\underline{\theta}}^{\underline{\theta}} q(\theta) dF(\theta|h)$ be the total quantity for agents in group h. The agent's payoff becomes $U(q, \theta, Q_h) - T(q)$.

Following Sundarajan (2004), agents' consumptions depend on the interactions between q and Q_h in their utility $U(q, \theta, Q_h)$. Using this model, Chen and Luo (2012) study network effects within seven phone books covering Toronto (Canada), where Q_h is the total amount of advertising in phone book h. There is a tariff $T_h(\cdot)$ for each phone book in their application. With the functional form $U(q, \theta, Q_h) = \theta[U_0(q) + W(q, Q_h)]$, data from the smallest phone book identify $U_0(q)$ and $f(\theta|h)$ for this phone book under the normalization W(q, Q) = 0 by the argument of Section II. Data from other phone books identify $W(\cdot, Q_h)$ and $f(\theta|h)$ for these phone books. Our estimation procedure applies, and their empirical results show some positive network effects. In the case of network effects within headings in the same phone book, the model is more complex because of a single tariff. This extension is under investigation.

B. Multiproduct Firm or Continuous Product Attributes

A firm can be multiproduct or a differentiated product can be viewed as a bundle of multiple continuous attributes. Luo et al. (2017) extend the current methodology to analyze nonlinear pricing with two goods, namely, voice and short message services in the cellular service industry. Every consumer has two types, one for each good. In an incomplete information setting, the major difficulty with multiple types is the multidimensional screening problem leading to pooling at equilibrium. Armstrong (1996) solves the problem with an optimal cost-based tariff T(c), where c is the cost for producing the product attributes. The basic idea is to exploit an aggregate type $h \equiv h(\cdot, ..., \cdot)$. Under some assumptions, the model provides two FOCs similar to (5) and (6) involving the marginal cost-based indirect utility $V_{0c}(c)$ and the aggregate type distribution $\Phi(\cdot)$. The model then defines two one-to-one mappings, one between h and c and the other between c and the payment t, where c is unobserved. We introduce unobserved product heterogeneity as a multiplicator on the variable cost leading to a transformation model. We show that the variable cost function and the fixed cost are identified. Once the cost is identified, our identification argument applies to identify the indirect costbased utility $V_0(\cdot)$ and the aggregate type density $\phi(\cdot)$. An estimator in the spirit of Section III.C is used.

C. Bundling, Discrete Attributes, and Differentiated Products

Luo (2015) develops a model combining bundling and nonlinear pricing in telecommunications in which phone service is measured continuously and internet is offered through J options. The model endogenizes bundling decisions based on utility complementarities, cost efficiency, and

correlation of types within individuals. Luo solves the multidimensional screening problem by partitioning the types space in the spirit of Rochet and Stole (2003). The consumer's utility is the sum of an intrinsic utility for phone $\theta U_0(Q)$ and a complementarity utility for phone and internet $V_j(Q)$, where j indexes the internet options, Q the voice quantity, and θ the voice type distributed as $F_j(\cdot)$. Under a parametric specification of the cost function, the model gives two FOCs similar to (5) and (6) for each internet level j. Using separately data from phone-only users, bundle users, and internet-only users, the author exploits these FOCs to identify the model primitives. The model can be estimated using the estimator of Section III.B.

Luo et al. (2015) extend this framework to differentiated products with several discrete and continuous attributes. Their model combines features of Armstrong (1996) and Luo (2015) with endogenization of product attributes. It leads to FOCs similar to (5) and (6) for each discrete attribute. In addition, the paper investigates how data from multiple markets and variations in exogenous variables can identify further the utility function and in particular the aggregate type function. This paper points to a general framework for the analysis of nonlinear pricing data with differentiated products.

D. Competition

A last extension concerns competition among firms, which is a challenging problem in nonlinear pricing, because of multidimensional screening as consumers are characterized by a vector of types whose dimension is the number of products/firms. A recent theoretical literature addresses nonlinear pricing and oligopoly with different results depending on modeling assumptions (see, e.g., Armstrong and Vickers 2001; Rochet and Stole 2003; Bonatti 2011). Considering the latter paper, which relates to the pure characteristics demand model in Berry and Pakes (2007), the consumer's payoff is $\theta_i Q_j - T_j(Q_j)$, where j = 1, ..., J indexes firms and $\theta = (\theta_1, \dots, \theta_I) \sim F(\cdot, \dots, \cdot)$ characterizes the consumer's tastes for products (Q_1, \ldots, Q_I) . The firm's maximization problem is similar to (4), where the integration domain is now the set $\Theta_i \subseteq \mathbb{R}^J$ of θ -consumers choosing the product j. Using Martimort and Stole (2002), the firm's problem reduces to a direct mechanism $[T_i(\theta_i), Q_i(\theta_i)]$. The IC and IR constraints are similar to (2) and (3). Bonatti (2011) derives the FOCs with either independent or correlated types and obtains an FOC similar to (5), where the distortion takes a complex form. The θ -consumer receives some rent, which determines all the consumers' rents, including those consuming the other products and hence the firms' market shares. Using an argument as in Section II.B, the joint distribution $F(\theta)$ is identified from the IC constraint while the equivalent of (5) leads to a second-order differential equation in the cost $C(\cdot)$. Estimation of the model is under investigation.

Despite all the difficulties associated with multidimensional screening in these extensions, the basic ideas of our methodology remain with some appropriate adjustments. Finally, the nonlinear pricing model contains the key ingredients to analyze contract data under incomplete information such as in labor, retailing, and regulation of pollution or utilities. Thus, our methodology opens several avenues for future empirical research.

Appendix

A. Proofs of Section II

Let S be the set of structures $[U_0(\cdot), F(\cdot), C(\cdot)]$ satisfying A1.

Lemma 1. Consider a structure $S = [U_0(\cdot), F(\cdot), C(\cdot)] \in \mathcal{S}$. Define another structure $\tilde{S} = [\tilde{U}_0(\cdot), \tilde{F}(\cdot), C(\cdot)]$, where $\tilde{U}_0(\cdot) = (1/\alpha)U_0(\cdot)$ and $\tilde{F}(\cdot) = F(\cdot/\alpha)$ for some $\alpha > 0$. Thus, $\tilde{S} \in \mathcal{S}$ and the two structures S and \tilde{S} lead to the same set of observables $[T(\cdot), G^{\mathbb{Q}^*}(\cdot)]$; that is, the two structures are observationally equivalent.

Proof of Lemma 1

Let $\tilde{\theta} = \alpha \theta$, which is distributed as $\tilde{F}(\cdot)$ on $[\underline{\tilde{\theta}}, \overline{\tilde{\theta}}] = [\alpha \underline{\theta}, \alpha \overline{\theta}]$. Let $\tilde{T}(\cdot) \equiv T(\cdot)$, $\tilde{Q}(\cdot) \equiv Q(\cdot/\alpha)$, and $\tilde{\theta}^* = \alpha \theta^*$. First we show that $\tilde{T}(\cdot)$, $\tilde{Q}(\cdot)$, and $\tilde{\theta}^*$ satisfy the necessary conditions (5), (6), and (7). We then show that $G^{\tilde{Q}^*}(\cdot) = G^{Q^*}(\cdot)$, where $G^{\tilde{Q}^*}(\cdot)$ is the truncated distribution of \tilde{Q} . Hence, the observables $[\tilde{T}(\cdot), G^{\tilde{Q}^*}(\cdot)]$ generated by the structure \tilde{S} are the same observables $[T(\cdot), G^{Q_*}(\cdot)]$ generated by the structure S. Finally, we show $\tilde{S} \in S$.

To show $\tilde{T}_Q(\tilde{Q}(\tilde{\theta})) = \tilde{\theta}'\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta}))$ for all $\tilde{\theta} \in (\tilde{\theta}^*, \tilde{\tilde{\theta}}]$, we rewrite this equation using the definition of $\tilde{T}(\cdot)$, $\tilde{U}_0(\cdot)$, and $\tilde{Q}(\cdot)$. This gives $T_Q(Q(\tilde{\theta}/\alpha)) = (\tilde{\theta}/\alpha)U_{0Q}(Q(\tilde{\theta}/\alpha))$ for all $\tilde{\theta} \in (\tilde{\theta}^*, \tilde{\tilde{\theta}}]$, which is true because of (6) with $\theta = \tilde{\theta}/\alpha \in [\theta^*, \tilde{\theta}]$. To show

$$\tilde{\theta}\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta})) = C_Q(\tilde{Q}(\tilde{\theta})) + \{[1 - \tilde{F}(\tilde{\theta})]/\tilde{f}(\tilde{\theta})\}\tilde{U}_{0Q}(\tilde{Q}(\tilde{\theta}))$$

for all $\tilde{\theta} \in (\tilde{\theta}^*, \tilde{\theta}]$, we rewrite this equation using the definition of $\tilde{U}_0(\cdot)$, $\tilde{Q}(\cdot)$, and $\tilde{F}(\cdot)$:

$$\frac{\tilde{\theta}}{\alpha} U_{0Q}(Q(\tilde{\theta}/\alpha)) = C_Q(Q(\tilde{\theta}/\alpha)) + \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)} U_{0Q}(Q(\tilde{\theta}/\alpha))$$

for all $\tilde{\theta} \in (\tilde{\theta}^*, \tilde{\theta}]$. The above equation holds for all $\theta = \tilde{\theta}/\alpha \in (\theta^*, \bar{\theta}]$ in view of (5). Regarding (7), we follow the same steps and obtain the equivalent of (7) with $\tilde{\theta}^*/\alpha$ for the argument of $Q(\cdot)$. This equation holds for $\tilde{\theta}^*/\alpha = \theta^*$ in view of (7).

Next, we show that the observables coincide. Since $\tilde{T}(\cdot) = T(\cdot)$, it suffices to show $G^{\tilde{Q}^*}(\cdot) = G^{Q^*}(\cdot)$. Namely,

$$\begin{split} G^{\tilde{Q}*}(y) &= \Pr[\tilde{Q}(\tilde{\theta}) \leq y | \tilde{Q}(\tilde{\theta}) > \tilde{Q}(\tilde{\theta}^*)] \\ &= \Pr[\tilde{\theta} \leq \tilde{Q}^{-1}(y) | \tilde{\theta} > \tilde{Q}^{-1}(Q(\tilde{\theta}^*/\alpha))] \\ &= \Pr[\alpha \theta \leq \alpha Q^{-1}(y) | \alpha \theta > \alpha Q^{-1}(Q(\theta^*))] \\ &= \Pr[\theta \leq Q^{-1}(y) | \theta > Q^{-1}(Q(\theta^*))] \\ &= \Pr[Q(\theta) \leq y | Q(\theta) > Q(\theta^*)] = G^{Q^*}(y), \end{split}$$

using the monotonicity of $\tilde{Q}(\cdot)$ and $Q(\cdot)$.

Finally, we verify that the structure \hat{S} belongs to S. Assumption A1-i, iii is trivially satisfied. Regarding A1-ii, we have

$$\tilde{\theta} - \frac{1 - \tilde{F}(\tilde{\theta})}{\tilde{f}(\tilde{\theta})} = \tilde{\theta} - \frac{1 - F(\tilde{\theta}/\alpha)}{(1/\alpha)f(\tilde{\theta}/\alpha)} = \alpha \left[\frac{\tilde{\theta}}{\alpha} - \frac{1 - F(\tilde{\theta}/\alpha)}{f(\tilde{\theta}/\alpha)} \right],$$

which is strictly increasing in $\tilde{\theta}/\alpha$ and hence in $\tilde{\theta}$. QED

Proof of Proposition 2

In view of the discussion in the text, it suffices to establish the identification of κ . From (7), we obtain

$$\begin{split} \kappa &= \theta^* U_0(Q(\theta^*)) - \gamma Q(\theta^*) - \frac{1 - F(\theta^*)}{f(\theta^*)} U_0(Q(\theta^*)) \\ &= U_0(\underline{Q}) \left[\theta^* - \frac{1 - F(\theta^*)}{f(\theta^*)} \right] - \gamma \underline{Q} \\ &= \gamma \frac{U_0(\underline{Q})}{U_{0Q}(\underline{Q})} - \gamma \underline{Q} = \gamma \left(\frac{T(\underline{Q})}{T_Q(\underline{Q})} - \underline{Q} \right), \end{split}$$

where the third equality is obtained from (5) evaluated at θ^* and the fourth equality exploits the boundary condition $\theta^* U_0(\underline{Q}) = T(\underline{Q})$ and (6) evaluated at $\theta^* = 1$ by B2. QED

Proof of Lemma 2

To prove necessity, we note that condition ii holds and $Q_{\theta}(\cdot) > 0$ on $[\theta^*, \bar{\theta}]$ from proposition 1. From the latter and $G^{Q^*}(Q(\theta)) = [F(\theta) - F(\theta^*)]/[1 - F(\theta^*)]$ for $\theta \in [\theta^*, \bar{\theta}]$, we have $g^{Q^*}(Q(\theta))Q_{\theta}(\theta) = f(\theta)/[1 - F(\theta^*)]$. Thus $G^{Q^*}(\cdot)$ satisfies condition i as $Q(\theta)$ is once differentiable from (5). In addition, $\bar{Q} = Q(\bar{\theta}) < \infty$ because $\bar{\theta} < \infty$. In the case of exclusion, condition iii follows proposition 2 since the fixed cost κ is nonnegative while γ is strictly positive from Al-iii. A similar argument applies to the case of no exclusion since the upper bound for κ must be nonnegative.

To prove sufficiency, we first construct a structure $[U_0(\cdot), F^*(\cdot), \kappa, \gamma]$ from $[T(\cdot), G^{Q^*}(\cdot)]$ satisfying conditions i–iii. We next show that this structure satisfies A1. The first part follows the identification argument of proposition 1. Namely, letting $\gamma = T_Q(\bar{Q}) > 0$ by proposition 2 and condition ii, we use (9) and B2 to construct

the quantile $\theta(\alpha)$ for every $\alpha \in [0, 1]$ of the truncated type distribution $F^*(\cdot)$. In particular, $\bar{\theta} = \theta(1) < \infty$ as (9) is well defined at $\alpha = 1$. This is so because

$$\begin{split} \lim_{\alpha\uparrow 1} [T_{\mathcal{Q}}(Q(\alpha)) - \gamma)]/(1 - \alpha) &= \lim_{\alpha\uparrow 1} [T_{\mathcal{Q}}(Q(\alpha)) - T_{\mathcal{Q}}(\bar{Q})]/(1 - \alpha) \\ &= -T_{00}(\bar{Q})Q_{\alpha}(1) < \infty, \end{split}$$

where $Q_{\alpha}(1) = 1/g^{Q^*}(\bar{Q}) < \infty$. This defines $F^*(\cdot)$ on $[\theta^*, \bar{\theta}]$. Thus using the FOC $T_Q(Q(\alpha)) = \theta(\alpha)U_{0Q}(Q(\alpha))$ gives $U_{0Q}(\cdot)$ on $[\underline{Q}, \bar{Q}]$. Using the boundary condition $U_0(\underline{Q}) = T(\underline{Q})$, we obtain the utility $U_0(\cdot)$ by integration. Finally, if there is exclusion, following proposition 2 we define the fixed cost κ as $T_Q(\bar{Q})\{[T(\underline{Q})/T_Q(\underline{Q})] - \underline{Q}\} \ge 0$ by condition iii. If there is no exclusion, we can choose any nonnegative κ bounded above by the latter value. Note that this structure satisfies the FOCs (5) and (6).

We now check that this structure $[U_0(\cdot), F^*(\cdot), \kappa, \gamma]$ satisfies A1. From (9), we have $\theta(\alpha) > 0$ for $\alpha \in [0,1]$ because $T(\cdot)$ is strictly concave. Moreover, from (8), $\theta_{\alpha}(\cdot)$ is strictly positive and continuously differentiable on [0,1] because $T_Q(\cdot)$ is strictly positive, decreasing, and continuously differentiable by condition ii and $Q(\cdot)$ is strictly increasing and continuously differentiable by condition i. In particular, $\theta_{\alpha}(1) = -\bar{\theta} T_{QQ}(\bar{Q}) Q_{\alpha}(1) / T_Q(\bar{Q})$, which is finite and strictly positive. Hence, $F^*(\cdot)$ is twice continuously differentiable with compact support $[\theta^*, \bar{\theta}] \subseteq \mathbb{R}_{++}$ and density $f^*(\theta(\alpha)) = 1/\theta_{\alpha}(\alpha)$, which is strictly positive on [0,1]. In particular, $f^*(\bar{\theta})$ is finite and strictly positive. Thus, $F^*(\cdot)$ satisfies A1-ii except for the condition on the hazard rate verified later.

Turning to the utility, we have $U_{0Q}(Q(\alpha)) = T_Q(Q(\alpha))/\theta(\alpha)$ showing that $U_{0Q}(\cdot) > 0$ on $[\theta^*, \bar{\theta}]$. Because of the boundary condition $U_0(Q) = T(Q) \ge 0$, it follows that $U_0(\cdot)$ is nonnegative on $[\underline{Q}, \bar{Q}]$. In addition, $U_{0Q}(\cdot)$ is once continuously differentiable as $T_Q(\cdot)$, $Q(\cdot)$, and $\theta(\cdot)$ are continuously differentiable. Taking the derivative with respect to α gives

$$T_{OO}(Q(\alpha))Q_{\alpha}(\alpha) - \theta_{\alpha}(\alpha)U_{OO}(Q(\alpha)) = \theta(\alpha)U_{OO}(Q(\alpha))Q_{\alpha}(\alpha),$$

showing that $U_{0QQ}(\cdot) < 0$ using $T_{QQ}(\cdot) < 0$, $Q_{\alpha}(\cdot) > 0$, $\theta_{\alpha}(\cdot) > 0$, $U_{0Q}(\cdot) > 0$, and $\theta(\cdot) > 0$. Thus, $U_0(\cdot)$ satisfies A1-i.

It remains to show that $\rho(\theta) \equiv \theta - \{[1 - F^*(\theta)]/f^*(\theta)\}$ is strictly positive and increasing. From (5), we have $\rho(\theta)U_{0Q}(Q(\theta)) = T_Q(\bar{Q}) > 0$ showing $\rho(\theta) > 0$ on $[\theta^*, \bar{\theta}]$ since $U_{0Q}(\cdot) > 0$ as shown above. Using this equation in quantiles and taking its derivative with respect to α gives

$$\rho_{\theta}(\theta(\alpha))\theta_{\alpha}(\alpha)U_{0O}(Q(\alpha)) + \rho(\theta(\alpha))U_{0OO}(Q(\alpha))Q_{\alpha}(\alpha) = 0,$$

showing $\rho_{\theta}(\cdot) > 0$ on $[\theta^*, \bar{\theta}]$ since $\theta_{\alpha}(\cdot) > 0$, $U_{0Q}(\cdot) > 0$, $\rho(\cdot) > 0$, $U_{0QQ}(\cdot) < 0$, and $Q_{\alpha}(\cdot) > 0$. Condition iii is trivially satisfied by a linear cost function with $\gamma = T_Q(\bar{Q}) > 0$. QED

B. Asymptotic Properties of Estimators in Section III.A

The next lemma establishes the strong consistency of $\hat{\kappa}$ and $\hat{\gamma}$ with rates of convergence faster than the parametric rate \sqrt{N} . It also provides their asymptotic distributions. These results follow from Galambos (1978) and the standard delta method. Let $\mathcal{E}(\lambda)$ denote the exponential distribution with parameter λ .

Lemma 3. Under A1, B1, and C1, as $N \to \infty$, we have

i.
$$\hat{\gamma} = \gamma + O_{\text{a.s.}}[(\log\log N)/N]$$
 and $\hat{\kappa} = \kappa + O_{\text{a.s.}}[(\log\log N)/N];$
ii. $N(\hat{\gamma} - \gamma) \xrightarrow{\mathcal{D}} \mathcal{E}(\lambda)$, where $\lambda = -g^{\mathcal{Q}^*}(\bar{\mathcal{Q}})/T_{\mathcal{Q}\mathcal{Q}}(\bar{\mathcal{Q}}) > 0$ and $N(\hat{\kappa} - \kappa) \xrightarrow{\mathcal{D}} \mathcal{E}(\lambda_1) + \mathcal{E}(\lambda_2)$, where $\mathcal{E}(\lambda_1)$ and $\mathcal{E}(\lambda_2)$ are mutually independent with

$$\lambda_1 \,=\, -\frac{g^{\mathcal{Q}^*}(\underline{\mathcal{Q}})\,T^2_{\mathcal{Q}}(\underline{\mathcal{Q}})}{T_{\mathcal{Q}}(\bar{\mathcal{Q}})\,T(\underline{\mathcal{Q}})\,T_{\mathcal{Q}\mathcal{Q}}(\underline{\mathcal{Q}})} > 0, \quad \lambda_2 \,=\, -\frac{g^{\mathcal{Q}^*}(\bar{\mathcal{Q}})\,T_{\mathcal{Q}}(\underline{\mathcal{Q}})}{T_{\mathcal{Q}\mathcal{Q}}(\bar{\mathcal{Q}})[T(\underline{\mathcal{Q}})-\underline{\mathcal{Q}}T_{\mathcal{Q}}(\underline{\mathcal{Q}})]} > 0.$$

We remark that λ_1 and λ_2 involve densities at boundaries with $g^{\mathcal{Q}^*}(\underline{Q})$ and $g^{\mathcal{Q}^*}(\bar{Q})$. Because of boundary effects, we follow Campo et al. (2011) and estimate $g^{\mathcal{Q}^*}(\bar{Q})$ and $g^{\mathcal{Q}^*}(\underline{Q})$ by one-sided kernel density estimators $\hat{g}^{\mathcal{Q}^*}(\cdot)$ evaluated at Q_{\max} and Q_{\min} , respectively. Replacing \underline{Q} and \bar{Q} by their estimates Q_{\min} and Q_{\max} , we can use the asymptotic distributions in ii to construct confidence intervals for the marginal cost γ and the fixed cost κ .

We now derive the asymptotic properties of $\hat{f}^*(\cdot)$ and $\hat{U}_{0Q}(\cdot)$. We proceed in two steps. In a first step, we study the properties of the quantile estimator $\hat{\theta}(\cdot)$ and its derivative $\hat{\theta}_{\alpha}(\cdot)$. This is done in lemma 4 in appendix B. Following the empirical process literature (see, e.g., Andrews 1994), we view $\hat{\theta}(\cdot)$ and $\hat{\theta}_{\alpha}(\cdot)$ as random elements in the space $\ell^{\infty}[0, \alpha_{\dagger}]$ of bounded functions on $[0, \alpha_{\dagger}]$ for any $\alpha_{\dagger} \in (0, 1)$. We equip the space $\ell^{\infty}[0, \alpha_{\dagger}]$ with its uniform metric $\|\psi_{1} - \psi_{2}\|_{\dagger} = \sup_{\alpha \in [0, \alpha_{\dagger}]} |\psi_{1}(\alpha) - \psi_{2}(\alpha)|$. Weak convergence on $\ell^{\infty}[0, \alpha_{\dagger}]$ is denoted by " \Rightarrow ."²² We then provide the uniform consistency and asymptotic distributions of the estimators (12) of $f^{*}(\cdot)$ and $U_{0Q}(\cdot)$ evaluated at their quantiles $\theta(\cdot)$ and $Q(\cdot)$, respectively. See (B.1) and (B.2) in appendix B. In the second step, we study the estimators $\hat{f}^{*}(\cdot)$ and $\hat{U}_{0Q}(\cdot)$ in (13) as compositions of estimated functions. The next proposition gives their asymptotic properties on the subsets $[\theta^{*}, \theta_{\dagger}]$ and $[Q, Q_{\dagger}]$, respectively, where $\theta_{\dagger} \in (\theta^{*}, \bar{\theta})$ and $Q_{\dagger} \in (Q, \bar{Q})$. For instance, $\theta_{\dagger} = \theta(\alpha_{\dagger})$ and $Q_{\dagger} = Q(\alpha_{\dagger})$ with $0 < \alpha_{\dagger} < 1$. Let $\ell^{\infty}[\theta^{*}, \theta_{\dagger}]$ and $\bar{\ell}^{\infty}[Q, Q_{\dagger}]$ denote the space of bounded functions on $[\theta^{*}, \theta_{\dagger}]$ and $[Q, Q_{\dagger}]$ equipped with their uniform metric $\|\cdot\|_{\dagger}$, respectively.

PROPOSITION 3. Under A1, B1, B2, and C1, for any $(\theta_{\dagger}, Q_{\dagger}) \in (\theta^*, \overline{\theta}) \times (\underline{Q}, \overline{Q})$, as $N \to \infty$, we have

i.
$$\|\hat{f}^*(\cdot) - f^*(\cdot)\|_{\dagger} \stackrel{\text{a.s.}}{\longrightarrow} 0$$
 and $\|\hat{U}_{0Q}(\cdot) - U_{0Q}(\cdot)\|_{\dagger} \stackrel{\text{a.s.}}{\longrightarrow} 0$;
ii. as random functions in $\ell^{\infty}[\theta^*, \theta_{\dagger}]$ and $\ell^{\infty}[\underline{Q}, Q_{\dagger}]$,

$$\sqrt{N} [\hat{f}^*(\theta) - f^*(\theta)] \Rightarrow -\gamma [f^*(\theta) + \theta f_{\theta}^*(\theta)] \mathcal{Z}[F^*(\theta)] + \frac{H_{\theta}(\theta)}{H(\theta)} \mathcal{B}_{F^*}(\theta)$$

$$\sqrt{N} [\hat{U}_{0Q}(Q) - U_{0Q}(Q)] \Rightarrow -U_{0Q}(Q) \left[\gamma \mathcal{Z}[G^{Q^*}(Q)] + \frac{T_Q(Q) - \gamma}{T_Q(Q)} \frac{\mathcal{B}_{G^{Q^*}}(Q)}{1 - G^{Q^*}(Q)}\right],$$

The ratio $[T_{\mathcal{Q}}(\hat{\mathcal{Q}}(\alpha)) - \hat{\gamma}]/(1-\alpha)$ is zero for $\alpha \in [(N-1)/N,1)$. In contrast, (8) and proposition 2 show that, as $\alpha \to 1$, $[T_{\mathcal{Q}}(\mathcal{Q}(\alpha)) - \gamma]/(1-\alpha)$ converges to $\gamma/[\bar{\theta}f^*(\bar{\theta})]$, which is finite and positive. Thus we could improve our estimator around the upper boundary by imposing this restriction on $\hat{\mathcal{Q}}(\cdot)$. To keep it simple, we choose instead to derive the asymptotic properties on $[0,\alpha_t)$, where $\alpha_t \in (0,1)$.

²² Measurability issues are ignored hereafter. As usual, this can be addressed by considering outer probabilities as in van der Vaart (1998).

where $H(\theta) = [1 - F^*(\theta)]/[\theta f^*(\theta)]$, $\mathcal{B}_{F^*}(\theta) = \mathcal{B}_{G^{\mathbb{C}^*}}[Q(\theta)]$, and $\mathcal{Z}(\cdot)$ is a tight Gaussian process defined on $[0, \alpha_t]$ by

$$\mathcal{Z}(\cdot) = -\int_{\underline{Q}}^{Q(\cdot)} \frac{T_{QQ}(q)}{T_Q^2(q)} \frac{\mathcal{B}_{G^{Q^*}}(q)}{1 - G^{Q^*}(q)} dq, \tag{A1}$$

with $\mathcal{B}_{G^{Q^*}}(\cdot)$ denoting the G^{Q^*} -Brownian bridge on $[Q, \bar{Q}]^{2^3}$

The first part establishes the uniform almost sure convergence of $\hat{f}^*(\cdot)$ and $\hat{U}_{0Q}(\cdot)$ on subsets $[\theta^*, \theta_\dagger] \subseteq [\theta^*, \bar{\theta})$ and $[\underline{Q}, Q_\dagger] \subseteq [\underline{Q}, \bar{Q})$, respectively. The second part gives the asymptotic distributions of these estimators. It is worth noting that their rates of convergence are the parametric rate \sqrt{N} . This contrasts with the previous literature on estimation of incomplete information models such as auctions (see, e.g., Guerre et al. 2000).

We remark that $\mathcal{Z}[G^{Q^*}(Q)]$ is given by the right-hand side of (A1) with $Q(\cdot)$ replaced by Q since $Q[G^{Q^*}(Q)] = Q$, while $\mathcal{Z}[F^*(\theta)]$ is given by

$$\mathcal{Z}[F^*(\theta)] = -\frac{1}{\gamma} \int_{\theta^*}^{\theta} H_{\theta}(x) \frac{\mathcal{B}_{F^*}(x)}{1 - F^*(x)} dx. \tag{A2}$$

The proof of the latter is given in appendix B. Using proposition 3-ii, we can derive the asymptotic distribution of $\hat{U}_0(Q) = T(Q_{\min}) + \int_{Q_{\min}}^{Q} \hat{U}_{0Q}(q) dq$. Because Q_{\min} is N-consistent, we obtain

$$\sqrt{N}[\hat{U}_0(Q) - U_0(Q)] \Rightarrow -\int_{\underline{Q}}^{\underline{Q}} U_{0Q}(q) \left\{ \gamma \mathcal{Z}[G^{Q^*}(q)] + \frac{T_Q(q) - \gamma}{T_Q(q)} \frac{\mathcal{B}_{G^{Q^*}}(q)}{1 - G^{Q^*}(q)} \right\} dq$$

uniformly in $Q \in [Q, Q_{\dagger}]$.

In practice, we use such asymptotic distributions to conduct large-sample hypothesis tests and construct pointwise or uniform confidence intervals provided that the asymptotic variances can be estimated consistently. Since we are mainly interested in the primitives $f^*(\cdot)$ and $U_{0Q}(\cdot)$, using proposition 3-ii, we have

$$\begin{split} & \sqrt{N}[\hat{f}^*(\theta) - f^*(\theta)] \xrightarrow{D} \mathcal{N}(0, V_{f^*}(\theta)), \\ & \sqrt{N}[\hat{U}_{0Q}(Q) - U_{0Q}(Q)] \xrightarrow{D} \mathcal{N}(0, V_{U_{0Q}}(Q)) \end{split}$$

for each $\theta \in [\theta^*, \theta_{\dagger}]$ and $Q \in [\underline{Q}, Q_{\dagger}]$, respectively. Appendix B provides detailed computations of the variances $V_f^*(\theta)$ and $V_{U_{bq}}(Q)$. It also discusses their consistent estimation.

²³ The G^{ϱ^*} -Brownian bridge on $[\underline{Q}, \bar{Q}]$ is the limit of the empirical process $(1/\sqrt{N})\Sigma_i\{1(Q_i \leq \cdot) - G^{\varrho^*}(\cdot)\}$ indexed by $[\underline{Q}, \bar{Q}]$; see (say) van der Vaart (1998, 266). It is a tight Gaussian process with mean 0 and covariance $G^{\varrho^*}(Q)[1-G^{\varrho^*}(Q')]$, where $\underline{Q} \leq Q \leq Q' \leq \bar{Q}$.

 $Q' \leq \overline{Q}$.

As for lemma 4 in app. B, at the lower boundary, $\hat{f}^*(\theta^*)$ and $\hat{U}_{0Q}(\underline{Q})$ converge at a faster rate, namely, at rate N, to their limits $f^*(\theta^*)$ and $U_{0Q}(\underline{Q}) = T_{\underline{Q}}(\underline{Q})$, respectively.

²⁵ Marmer and Shneyerov (2012) propose a quantile-based estimator using $v(\alpha) = b(\alpha) + [\alpha b_{\alpha}(\alpha)]/(I-1)$, where $v(\alpha)$ and $b(\alpha)$ are the quantiles of the unobserved private distribution and the observed bid distribution, respectively, while I denotes the number of bidders. Because the derivative $b_{\alpha}(\alpha)$ cannot be estimated at the parametric rate, their estimator cannot achieve \sqrt{N} -consistency.

Finally, regarding the estimators (15)–(18) based on payments, lemmas 3 and 4 as well as proposition 3 still hold, while the asymptotic variances can be written directly in terms of $G^{t^*}(\cdot)$. For instance, because t = T(Q), the Gaussian process $\mathcal{Z}(\cdot)$ on $[0, \alpha_t]$ can be written as

$$\mathcal{Z}(\cdot) = \int_{t}^{t(\cdot)} T_{tt}^{-1}(t) \frac{\mathcal{B}_{G^{t}}(t)}{1 - G^{t^{*}}(t)} dt$$
 (A3)

from the change-of-variable t = T(q) in (A1) so that $T_Q(q) = 1/T_t^{-1}(t)$, $T_{QQ}(q) = -T_t^{-1}(t)/T_t^{-1}(t)^3$, and $G^{Q^*}(q) = G^{P^*}(t)$.

C. Asymptotic Properties of Estimators in Section III.B

The payment density $g^{T^*}(\cdot)$ is strictly positive on its support $[\underline{t}, \overline{t}]$, with $0 < \underline{t} < \overline{t} < \infty$ since $t = T[Q(\theta)]$, $T_Q(\cdot) > 0$ on $[\underline{Q}, \overline{Q}]$, $Q_\theta(\cdot) > 0$ on $[\theta^*, \theta]$, and $f^*(\cdot) > 0$ on its support $[\theta^*, \overline{\theta}]$. Because $T^{-1}(\cdot; \widehat{\beta})$ and $T_t^{-1}(\cdot; \widehat{\beta})$ converge uniformly on $[\underline{t}, \overline{t}]$ to $T^{-1}(\cdot) = T^{-1}(\cdot; \beta)$ and $T_t^{-1}(\cdot) = T_t^{-1}(\cdot; \beta)$, respectively, consistency of our estimators (20)–(23) is straightforward. Thus we focus on their asymptotic distributions. Our derivation underscores the effects of estimating β in $T^{-1}(\cdot; \beta)$.

We begin with the estimators of the marginal cost γ and fixed cost κ .

Lemma 5. Under A1, B1, and C1, as $N \to \infty$ we have

i.
$$\sqrt{N}(\tilde{\gamma} - \gamma) = -\gamma^2 T_{\iota\beta}^{-1}(\bar{t}; \beta) \sqrt{N}(\hat{\beta} - \beta) + o_P(1)$$
 and
ii. $\sqrt{N}(\tilde{\kappa} - \kappa) = \gamma(\underline{t}T_{\iota\beta}^{-1}(\underline{t}; \beta) - T_{\beta}^{-1}(\underline{t}; \beta) - \kappa T_{\iota\beta}^{-1}(\bar{t}; \beta)) \sqrt{N}(\hat{\beta} - \beta) + o_P(1)$.

Since $\sqrt{N}(\hat{\beta} - \beta) \stackrel{D}{\longrightarrow} \mathcal{N} \sim \mathcal{N}(0, \Omega)$, lemma 5 implies that $\tilde{\gamma}$ and $\tilde{\kappa}$ are asymptotically normal. In contrast to lemma 3, their convergence rate is now \sqrt{N} instead of N. The reason is that $\hat{\beta}$ converges at a slower rate than the infeasible estimators $\hat{\gamma}$ and $\hat{\kappa}$ when the tariff is known. As usual, consistent estimation of the asymptotic variances of $\sqrt{N}(\tilde{\gamma} - \gamma)$ and $\sqrt{N}(\tilde{\kappa} - \kappa)$ is obtained by replacing γ , \underline{t} , \overline{t} , and β by their consistent estimators $\tilde{\gamma}$, t_{\min} , t_{\max} , and $\hat{\beta}$ upon consistent estimation of the asymptotic variance Ω of $\sqrt{N}(\hat{\beta} - \beta)$.

Next, we turn to $\tilde{f}^*(\cdot)$ and $\tilde{U}_{0Q}(\cdot)$. We first study the asymptotic properties of the estimators of the quantile $\theta(\cdot)$ and its derivative $\theta_{\alpha}(\cdot)$. See lemma 6 in appendix C, which shows the effects of estimating β relative to the infeasible estimators $\hat{\theta}(\cdot)$ and $\hat{\theta}_{\alpha}(\cdot)$. We now derive the asymptotic distributions of $\tilde{f}^*(\cdot)$ and $\tilde{U}_{0Q}(\cdot)$ on $[\theta^*, \theta_{\dagger}]$ and $[\underline{Q}, Q_{\dagger}]$, where $\theta_{\dagger} \in (\theta^*, \bar{\theta})$ and $Q_{\dagger} \in (\underline{Q}, \bar{Q})$.

Proposition 4. Under A1, B1, B2, and C1, for $\bar{\theta}_{\dagger} \in (\theta^*, \bar{\theta})$ and $Q_{\dagger} \in (\underline{Q}, \bar{Q})$, as $N \to \infty$ we have

$$\begin{split} \sqrt{N} [\tilde{f}^*(\theta) - f^*(\theta)] &\Rightarrow -\gamma \big[f^*(\theta) + \theta f_{\theta}^*(\theta) \big] \mathcal{Z}[F^*(\theta)] \\ &\quad + \frac{H_{\theta}(\theta)}{H(\theta)} \mathcal{B}_{F^*}(\theta) - \gamma b(\theta) \mathcal{N}, \\ \\ \sqrt{N} \big[\tilde{U}_{0Q}(Q) - U_{0Q}(Q) \big] &\Rightarrow -U_{0Q}(Q) \left\{ \gamma \mathcal{Z}[G^{Q^*}(Q)] \right. \\ &\quad + \frac{T_Q(Q;\beta) - \gamma}{T_O(Q;\beta)} \frac{\mathcal{B}_{G^{Q^*}}(Q)}{1 - G^{Q^*}(Q)} + c(Q) \mathcal{N} \right\}, \end{split}$$

as random functions in $\ell^{\infty}[\theta^*, \theta_{\dagger}]$ and $\ell^{\infty}[Q, Q_{\dagger}]$, respectively, where

$$\begin{split} b(\theta) &= [f^*(\theta) + \theta f_{\theta}^*(\theta)] I[F^*(\theta)] - \frac{f^*(\theta)}{H(\theta)} a[F^*(\theta)], \\ c(Q) &= \gamma I[G^{Q^*}(Q)] + \frac{T_{i\beta}^{-1}(T;\beta)}{T_t^{-1}(T;\beta)} + \frac{U_{0QQ}(Q)}{U_{0Q}(Q)} T_{\beta}^{-1}(T;\beta) \end{split}$$

are nonstochastic $(1 \times \dim \beta)$ vectors and $T = T(Q; \beta)$.

For each statement, the first two terms are identical to the limiting process of the infeasible estimator in proposition 3, while the third term arises from estimating β as $\sqrt{N}(\hat{\beta} - \beta) \stackrel{D}{\rightarrow} \mathcal{N} \sim \mathcal{N}(0, \Omega)$.

The limiting processes in proposition 4 are tight Gaussian processes with zero means and finite covariance functions. A difficulty in determining the latter is that the processes $\mathcal{Z}(\cdot)$ and $\mathcal{B}(\cdot)$ are not independent from the random vector \mathcal{N} as lemma C.1 shows. The reason is that $\mathcal{Z}[G^{Q^*}(\cdot)]$ and $\mathcal{B}_{G^{Q^*}}(\cdot)$ arise from $\{Q_i = T^{-1}(t_i); i = 1, ..., N\}$, while \mathcal{N} arises from $\hat{\beta}$, which depends on $\{(t_i, q_i); i = 1, ..., N\}$. Using proposition 4, we obtain

$$\sqrt{N} ig[ilde{f}^*(heta) - f^*(heta) ig] \overset{\scriptscriptstyle D}{ o} \mathcal{N} \Big(0, \omega_{f^*}^2(heta) \Big), \ \sqrt{N} ig[ilde{U}_{0Q}(Q) - U_{0Q}(Q) ig] \overset{\scriptscriptstyle D}{ o} \mathcal{N} \Big(0, \omega_{U_{b_Q}}^2(Q) \Big)$$

for each $\theta \in [\theta^*, \theta_\dagger]$ and $\theta \in [Q, Q_\dagger]$, respectively. Given a linear representation of $\sqrt{N}(\hat{\beta} - \beta)$, appendix C derives $\omega_{f*}^2(\theta)$ and $\omega_{\ell_{bq}}^2(Q)$ and discusses their consistent estimation.

D. Estimation of the Tariff Function

We partition the range $[\underline{t}, \overline{t}]$ into equally spaced K+1=5 bins of the form $[\tau_{k-1}, \tau_k)$ for $k=1, \ldots, 5$ with $\tau_0 = t_{\min}$ and $\tau_5 = t_{\max}$. Our basis functions $\psi_k(\cdot)$'s are

$$\psi_k(t) = \begin{cases} 0 & \text{if } t \in [\tau_0, \tau_{k-1}] \\ (t - \tau_{k-1})^3 / [6(\tau_k - \tau_{k-1})] & \text{if } t \in [\tau_{k-1}, \tau_k] \\ \{(t - \tau_{k+1})^3 / [6(\tau_k - \tau_{k+1})]\} + a_{1k}t + a_{0k} & \text{if } t \in [\tau_k, \tau_{k+1}] \\ a_{1k}t + a_{0k} & \text{if } t \in [\tau_{k+1}, \tau_{K+1}] \end{cases}$$

for k = 1, ..., K, where $a_{1k} = (\tau_{k+1} - \tau_{k-1})/2$ and $a_{0k} = [(\tau_k - \tau_{k-1})^2 - (\tau_k - \tau_{k+1})^2 - 3\tau_k(\tau_{k+1} - \tau_{k-1})]/6$, and

$$\begin{split} \psi_{K+1}(t) &= \begin{cases} 0 & \text{if } t \in [\tau_0, \tau_K] \\ (t - \tau_K)^3 & \text{if } t \in [\tau_K, \tau_{K+1}], \end{cases} \\ \psi_{K+2}(t) &= \begin{cases} \left\{ (\tau_1 - t)^3 / [6(\tau_1 - \tau_0)] \right\} + a_{1K+2}t + a_{0K+2} & \text{if } t \in [\tau_0, \tau_1] \\ a_{1K+2}t + a_{0K+2} & \text{if } t \in [\tau_1, \tau_{K+1}], \end{cases} \end{split}$$

where $a_{1K+2} = (\tau_1 - \tau_0)/2$ and $a_{0K+2} = (2\tau_0^2 - \tau_1^2 - \tau_0\tau_1)/6$.

We choose two normalizations at the lowest payment $t_{\min}=13.97$. Given (24), we have

$$T^{-1}(t_{\min}) = \beta_0 + \beta_1 t_{\min}, \quad T_t^{-1}(t_{\min}) = \beta_1.$$
 (A4)

We estimate $T^{-1}(t_{\min})$ and $T_t^{-1}(t_{\min})$ by a second-degree local polynomial regression of q on t using a Gaussian kernel and a standard bandwidth. We obtain $\hat{T}^{-1}(t_{\min}) = 195.2$ and $\hat{T}_t^{-1}(t_{\min}) = 46.11$, thereby providing $\hat{\beta}_0 = -448.96$ and $\hat{\beta}_1 = 46.11$ by solving (A4).

Hereafter $\beta = (\hat{\beta}_0, \hat{\beta}_1, \delta_1, \dots, \delta_6)$. We use Linton et al.'s (2008) minimum distance estimator to estimate $\delta = (\delta_1, \dots, \delta_6)$. For any fixed β , let $\log \hat{\epsilon}_i(\beta)$ be the ith residual of a nonparametric regression of $\log T^{-1}(t;\beta)$ on q, that is, $\log \hat{\epsilon}_i(\beta) = \log T^{-1}(t;\beta) - \log \hat{m}_0(q_i;\beta)$, where $\log \hat{m}_0(\cdot;\beta)$ is a kernel estimator of the regression $\mathbb{E}[\log T^{-1}(t;\beta)|q=\cdot]$. We use a triweight kernel $K(u) = (35/32)(1-u^2)^3 \mathbf{1}(|u| \le 1)$ and a standard bandwidth $h = 1.06\hat{\sigma}_q N^{-1/5} = 86.34$, where $\hat{\sigma}_q$ is the empirical standard deviation of the q_i 's. For any fixed β , let the empirical cumulative distribution functions

$$\hat{G}^q(q) = rac{1}{N} \sum_{i=1}^N \mathbf{1}(q_i \leq q), \quad \hat{F}^{\epsilon(eta)}(e) = rac{1}{N} \sum_{i=1}^N \mathbf{1}(\hat{\epsilon}_i(eta) \leq e),$$

$$\hat{F}^{q,\epsilon(eta)}(q,e) = rac{1}{N} \sum_{i=1}^N \mathbf{1}(q_i \leq q) \mathbf{1}(\hat{\epsilon}_i(eta) \leq e).$$

The estimator $\hat{\beta}$ is obtained through a grid search minimizing the criterion function

$$Q_{\!\scriptscriptstyle N}(\beta) \, = \, \frac{1}{N} \sum_{}^{N} \big[\hat{R}^{q,\epsilon(\beta)}(q_{\!\scriptscriptstyle i},\hat{\epsilon}_{\scriptscriptstyle i}(\beta)) \, - \, \hat{G}^{q}(q_{\!\scriptscriptstyle i}) \hat{R}^{\epsilon(\beta)}(\hat{\epsilon}_{\scriptscriptstyle i}(\beta)) \big]^2.$$

Given $\hat{\beta}$, an estimate of $m_0(\cdot)$ is obtained by a kernel regression of $\log T^{-1}(t_i; \hat{\beta})$ on q_i giving $\log \hat{m}_0(\cdot) = \hat{\mathbb{E}}[\log T^{-1}(t_i; \hat{\beta})|q=\cdot]$. The density of $\log \epsilon$ is then estimated by a kernel density estimator from the estimated residuals $\log \hat{\epsilon}_i(\hat{\beta})$ using the above kernel and bandwidth equal to 0.193 based on $\hat{\sigma}_{\epsilon}$.

The test of independence of q and ϵ is based on a model restriction that follows from the identification argument in Horowitz (1996). Specifically, letting $F_{\ell|q}(t|q) \equiv \phi(t|q)$, the transformation model (19) leads to $T_{\iota}^{-1}(t)/T^{-1}(t) = -[m_{0q}(q)/m_0(q)]s(t,q)$, where $s(t,q) \equiv \phi_{\iota}(t|q)/\phi_q(t|q)$. Evaluated at t_{\min} , we obtain

$$\left[m_{0q}(q)/m_0(q) \right] \, = \, -(46.11/195.2)[1/s(t_{\rm min},\,q)]$$

using the above normalizations. Moreover, integrating from t_{\min} to t gives

$$\log T^{-1}(t) = \log 195.2 - \left[m_{0q}(q) / m_0(q) \right] S(t, q)$$

= \log 195.2 + (46.11/195.2) \[S(t, q) / s(t_{\text{min}}, q) \],

where $S(t,q) \equiv \int_{t_{min}}^{t} s(u,q) du$. Thus the right-hand side should be independent of q, which is the main restriction imposed by the transformation model under the independence of ϵ and q. Chiappori et al. (2015) then develop a test of the null assumption of independence using subsampling to approximate the asymptotic distribution.

References

- Andrews, Donald W. K. 1994. "Empirical Process Methods in Econometrics." In Handbook of Econometrics, vol. 4, edited by Robert F. Engle and Daniel L. McFadden. Amsterdam: North-Holland.
- Armstrong, Mark. 1996. "Multiproduct Nonlinear Pricing." *Econometrica* 64 (1): 51–76.
- Armstrong, Mark, and John Vickers. 2001. "Competitive Price Discrimination." *Rand J. Econ.* 32 (4): 579–605.
- Athey, Susan, and Philip A. Haile. 2007. "Nonparametric Approaches to Auctions." In *Handbook of Econometrics*, vol. 6, edited by James J. Heckman and Edward E. Leamer. Amsterdam: North-Holland.
- Berry, Steven, and Ariel Pakes. 2007. "The Pure Characteristics Demand Model." *Internat. Econ. Rev.* 48 (4): 1193–1225.
- Bonatti, Alessandro. 2011. "Brand-Specific Tastes for Quality." *Internat. J. Indus. Org.* 29 (5): 562–75.
- Borenstein, Severin. 1991. "Selling Costs and Switching Costs: Explaining Retail Gasoline Margins." *Rand J. Econ.* 22 (3): 354–69.
- Borenstein, Severin, and Nancy L. Rose. 1994. "Competition and Price Dispersion in the U.S. Airline Industry." *J.P.E.* 102 (4): 653–83.
- Busse, Meghan, and Marc Rysman. 2005. "Competition and Price Discrimination in Yellow Pages Advertising." *Rand J. Econ.* 36 (2): 378–90.
- Campo, Sandra, Emmanuel Guerre, Isabelle Perrigne, and Quang Vuong. 2011. "Semiparametric Estimation of First-Price Auctions with Risk Averse Bidders." *Rev. Econ. Studies* 78 (1): 112–47.
- Chen, Liang, and Yao Luo. 2012. "Nonlinear Pricing with Network Effects in Yellow Pages." Manuscript, Dept. Econ., Univ. Toronto.
- Chen, Songnian. 2002. "Rank Estimation of Transformation Models." *Econometrica* 70 (4): 1683–97.
- Chesher, Andrew. 2003. "Identification in Nonseparable Models." *Econometrica* 71 (5): 1405–41.
- Chiappori, Pierre-André, Ivana Komunjer, and Dennis Kristensen. 2015. "Non-parametric Identification and Estimation of Transformation Models." J. Econometrics 188 (1): 22–39.
- Cohen, Andrew. 2008. "Package Size and Price Discrimination in Paper Towels." Internat. J. Indus. Org. 26 (2): 502–16.
- Crawford, Gregory S., and Matthew Shum. 2007. "Monopoly Quality Degradation and Regulation in Cable Television." *J. Law and Econ.* 50 (1): 181–219.
- Devroye, Luc, and Gary L. Wise. 1980. "Detection of Abnormal Behavior via Non-parametric Estimation of the Support." SIAM J. Appl. Math. 38 (3): 480–88.
- D'Haultfoeuille, Xavier, and Philippe Février. 2007. "Identification and Estimation of Incentive Problems: Adverse Selection." Working Paper no. 2007-21, Center Res. Econ. and Statis., Paris.
- ——. 2016. "The Provision of Wage Incentives: A Structural Estimation Using Contract Variation." Manuscript, Center Res. Econ. and Statis., Paris.
- Dole, David. 1999. "CoSmo: A Constrained Scatterplot Smoother for Estimating Convex, Monotonic Transformations." J. Bus. and Econ. Statis. 17 (4): 444–55.
- Economides, Nicholas, Katja Seim, and V. Brian Viard. 2008. "Quantifying the Benefits of Entry into Local Phone Service." *Rand J. Econ.* 39 (3): 699–730.
- Ekeland, Ivar, James J. Heckman, and Lars Nesheim. 2004. "Identification and Estimation of Hedonic Models." *J.P.E.* 112 (February): 60–109.

- Galambos, Janos. 1978. The Asymptotic Theory of Extreme Order Statistics. New York: Wiley.
- Gandhí, Amit, Salvador Navarro, and David A. Rivers. 2013. "On the Identification of Production Functions: How Heterogeneous Is Productivity?" Manuscript, Univ. Pennsylvania.
- Grubb, Michael D., and Matthew Osborne. 2015. "Cellular Service Demand: Biased Beliefs, Learning and Bill Shock." A.E.R. 105 (1): 234–71.
- Guerre, Emmanuel, Isabelle Perrigne, and Quang Vuong. 2000. "Optimal Nonparametric Estimation of First-Price Auctions." *Econometrica* 68 (3): 525–74.
- 2009. "Nonparametric Identification of Risk Aversion in First-Price Sealed-Bid Auctions under Exclusion Restrictions." *Econometrica* 77 (4): 1193–1297.
- Heckman, James J., Rosa L. Matzkin, and Lars Nesheim. 2010. "Nonparametric Identification and Estimation of Nonadditive Hedonic Models." *Econometrica* 78 (5): 1569–91.
- Horowitz, Joel L. 1996. "Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable." *Econometrica* 64 (1): 103–37.
- Ito, Koichiro. 2014. "Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing." *A.E.R.* 104 (2): 537–63.
- Ivaldi, Marc, and David Martimort. 1994. "Competition under Nonlinear Pricing." *Annales d'Economie et de Statistique* 34:71–114.
- Khan, Shakeeb, Youngki Shin, and Elie Tamer. 2011. "Heteroscedastic Transformation Models with Covariate Dependent Censoring." *J. Bus. Econ. and Statis.* 29 (1): 40–48.
- Laffont, Jean-Jacques, and Quang Vuong. 1996. "Structural Analysis of Auction Data." A.E.R. Papers and Proc. 86 (2): 414–20.
- Leslie, Phillip. 2004. "Price Discrimination in Broadway Theatre." *Rand J. Econ.* 35 (3): 520–41.
- Lewis, Tracy R., and David E. M. Sappington. 1989. "Countervailing Incentives in Agency Problems." *J. Econ. Theory* 49 (2): 294–313.
- Linton, Oliver, Stefan Sperlich, and Ingrid Van Keilegom. 2008. "Estimation of a Semiparametric Transformation Model." *Ann. Statis.* 36 (2): 686–718.
- Lott, John R., Jr., and Russell D. Roberts. 1991. "A Guide to the Pitfalls of Identifying Price Discrimination." *Econ. Inquiry* 29 (1): 14–23.
- Luo, Yao. 2015. "Bundling and Nonlinear Pricing in Telecommunications." Manuscript, Dept. Econ., Univ. Toronto.
- Luo, Yao, Isabelle Perrigne, and Quang Vuong. 2015. "A General Framework for Nonlinear Pricing Data." Manuscript, New York Univ.
- 2017. "Multiproduct Nonlinear Pricing: Mobile Voice Service and SMS." Manuscript, New York Univ.
- Maggi, Giovanni, and Andres Rodriguez-Clare. 1995. "On Countervailing Incentives." *J. Econ. Theory* 66 (1): 238–63.
- Marmer, Vadim, and Artyom Shneyerov. 2012. "Quantile-Based Nonparametric Inference for First-Price Auctions." *J. Econometrics* 167 (2): 345–57.
- Martimort, David, and Lars Stole. 2002. "The Revelation and Delegation Principles in Common Agency Games." *Econometrica* 70 (4): 1659–73.
- Maskin, Eric, and John Riley. 1984. "Monopoly with Incomplete Information." *Rand J. Econ.* 15 (2): 171–96.
- Matzkin, Rosa L. 2003. "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica* 71 (5): 1339–75.

- McManus, Brian. 2007. "Nonlinear Pricing in an Oligopoly Market: The Case of Specialty Coffee." *Rand J. Econ.* 38 (2): 512–32.
- Meyer, Mary C. 2012. "Constrained Penalized Splines." *Canadian J. Statis.* 40 (1): 190–206.
- Miravete, Eugenio J. 2002. "Estimating Demand for Local Telephone Service with Asymmetric Information and Optional Calling Plans." *Rev. Econ. Studies* 69 (4): 943–71.
- Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." J. Econ. Theory 18 (2): 301–17.
- Oren, Shmuel S., Stephen A. Smith, and Robert B. Wilson. 1983. "Competitive Nonlinear Tariffs." *J. Econ. Theory* 29 (1): 49–71.
- Perrigne, Isabelle, and Quang Vuong. 2011. "Nonparametric Identification of a Contract Model with Adverse Selection and Moral Hazard." *Econometrica* 79 (5): 1499–1539.
- Riley, John G. 2012. Essential Microeconomics. Cambridge: Cambridge Univ. Press. Rochet, Jean-Charles, and Philippe Choné. 1998. "Ironing, Sweeping and Multidimensional Screening." Econometrica 66 (4): 783–826.
- Rochet, Jean-Charles, and Lars A. Stole. 2003. "The Economics of Multidimensional Screening." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky. Cambridge: Cambridge Univ. Press.
- Shepard, Andrea. 1991. "Price Discrimination and Retail Configuration." *J.P.E.* 99 (1): 30–53.
- Spence, Michael. 1977. "Nonlinear Prices and Welfare." J. Public Econ. 8 (1): 1–18.
- Stole, Lars A. 1995. "Nonlinear Pricing and Oligopoly." J. Econ. Management and Strategy 4 (4): 529–62.
- ——. 2007. "Price Discrimination and Competition." In *Handbook of Industrial Organization*, vol. 3, edited by Mark Armstrong and Robert Porter. Amsterdam: North-Holland.
- Sundararajan, Arun. 2004. "Nonlinear Pricing and Type-Dependent Network Effects." *Econ. Letters* 83 (1): 107–13.
- Tirole, Jean. 1988. The Theory of Industrial Organization. Cambridge, MA: MIT Press
- van der Vaart, A. 1998. *Asymptotic Statistics*. Cambridge: Cambridge Univ. Press. Wilson, Robert B. 1993. *Nonlinear Pricing*. Oxford: Oxford Univ. Press.
- Wong, Adam. 2012. "Menu Implementability and Practical Pricing Schemes." Manuscript, Lingnan Univ.
- Yao, Song, Ĉarl F. Mela, Jeongwen Chiang, and Yuxin Chen. 2012. "Determining Consumers' Discount Rates with Field Studies." *J. Marketing Res.* 49 (6): 822–41